

Thesis Title:

Development of a CONSORT Extension for  
Social and Psychological Interventions

Mr Sean Patrick Grant, MSc

Green Templeton College

Candidate Number: 238916

DPhil in Social Intervention Thesis

Trinity Term 2014



## Acknowledgements

“Whensoever thou wilt rejoice thyself, think and meditate upon those good parts and especial gifts, which thou hast observed in any of them that live with thee: as industry in one, in another modesty, in another bountifulness, in another some other thing. For nothing can so much rejoice thee, as the resemblances and parallels of several virtues, eminent in the dispositions of them that live with thee, especially when all at once, as it were, they represent themselves unto thee. See therefore, that thou have them always in a readiness.”

- Marcus Aurelius, *Meditations*

I have an endless list of family, friends, and mentors to thank for this thesis and for my Oxford experience. I will try to list them briefly below, but know that I am grateful for all those who have been in my life over the last four years.

I first and foremost would like to thank all participants in the CONSORT-SPI project—the trialists and previous guideline developers, those who took part in the Delphi process and consensus meeting, and the advisory board. Thank you for giving of your time and efforts in contributing to this important endeavor.

I thank with greatest sincerity my supervisor, Paul Montgomery. You took a chance on me from my first day at Oxford, and supported me through and through until my last day. Of all of our experiences and memories, I am most thankful for your lesson that work is only work, and to make sure that work is not all there is to life.

To Evan Mayo-Wilson, I am grateful for our growing relationship—from being your student, to your colleague, to your friend. I am looking forward to years more of collaboration.

To the other faculty at the Department of Social Policy and Intervention, thank you for your world-class instruction, and abundant opportunities to apply research to the “real world”.

To Mrs. B and the rest of the CEBI crew, thank you for countless nights at the Oak and elsewhere, and for making Oxford truly feel like a second home.

To my examiners, thank you for challenging me on every assumption and facet of my DPhil project and facilitating my thinking as an independent scholar.

To my fellow Graduate Research Students, thank you for constantly challenging me to grow intellectually through your brilliance and communal bond in the eternal struggle that is a DPhil.

To my GTC crew—LB, the Have, the Rowing Club, the Footballers, and most importantly the epic GTC Cover Band—I am forever thankful for the friendships that I have made that will last for a lifetime. You always have a couch in Los Angeles to crash on—and I expect the favor to be returned!

To my family, thank you for always believing in my dreams, scientific or otherwise. Never once have you doubted that I could pursue this vocation. I am so incredibly lucky to have you in my life.

Lastly, I'd like to thank Lisa. Through Skype calls, trips to Nebraska, and my endless anxiety, you have stayed by me and supported me through thick and thin. You've waited four years for me to finish this project and come home. I love you, and I can't wait to spend the rest of our years together.

*Development of a CONSORT Extension for  
Social and Psychological Interventions*

**Abstract**

**Background:** Defined by their mechanisms, social and psychological interventions are those interventions that work through mental processes and social phenomena. They are often complex and challenging to evaluate, so understanding randomised controlled trials (RCTs) of these interventions requires detailed reports of the interventions tested and the methods used to assess them. However, reports of these RCTs often omit important information. Poor reporting hinders critical appraisal and synthesis of RCTs in systematic reviews, thereby impeding the effective transfer of research evidence to policy and practice. The Consolidated Standards for Reporting Trials (CONSORT) Statement is a reporting guideline that has contributed to improvements in the quality of RCT manuscripts in journals publishing medical research. However, studies have shown persistent deficiencies in the reporting quality of social and psychological intervention trials. A new CONSORT extension for these interventions may be needed given their distinct and complex features. This DPhil thesis reports on a project to develop and disseminate an official CONSORT Extension for Social and Psychological Interventions: CONSORT-SPI.

**Structure:** Following a preface, this DPhil thesis includes eight chapters. Chapter 1 provides an overview of the conceptual rationale that prompted the CONSORT-SPI project. Chapter 2 details the project protocol, which consists of a five-phase methodology that follows current best practices for reporting guideline development and dissemination. Chapter 3 discusses systematic literature reviews to assess reporting guidelines for and the reporting quality of publications of social and psychological

intervention RCTs. Chapter 4 discusses an online, international Delphi process to generate a prioritised list of possible items to include in the CONSORT-SPI extension. Chapter 5 discusses a formal consensus meeting to select reporting items to add to or modify for the CONSORT-SPI Extension checklist. Chapter 6 involves drafts of the CONSORT-SPI checklist as well as a template for the Explanation and Elaboration (E&E) document providing detailed advice and examples of good reporting for each checklist item. These drafts have not yet been circulated to co-authors or other members of the project team; their purpose in this thesis is to give an indication of how previous project phases have led into initial prototypes of the checklist and E&E, which will undergo further development and revision by the project team before publication. Chapter 7 proposes a coordinated dissemination and implementation strategy informed by theoretical frameworks and tools used to guide the implementation of clinical guidelines and empirically-supported interventions. The final chapter summarises the information gained from the CONSORT-SPI project to date, assesses strengths and limitations of the project methodology, and discusses implications for future research.

**Conclusion:** A CONSORT-SPI Extension could improve the reporting quality of social and psychological intervention RCTs. This extension could also facilitate better critical appraisal of this body of research and its use in evidence-based decision-making. With successful dissemination and implementation, the guideline will hopefully contribute to the improvement of intervention evaluations—as well as the methodology underpinning these studies—within the social and behavioural sciences.

## Table of Contents

### *Preface to DPhil Thesis*

1. What are Social and Psychological Interventions?	1
2. Randomised Controlled Trials of Social and Psychological Interventions	3
3. Reporting Standards and Reporting Guidelines	4
3.1 The CONSORT Statement	5
3.2 A CONSORT Extension for Social and Psychological Interventions?	5
4. Overview of Thesis Structure	6
<b>Figure 1. Workflow for Thesis on CONSORT-SPI Project</b>	<b>7</b>
4.1 Chapter 1 of the Thesis: Conceptual Overview	8
4.2 Chapter 2 of the Thesis: Project Protocol	9
4.3 Chapter 3 of the Thesis: Systematic Reviews	9
4.4 Chapter 4 of the Thesis: Delphi Process	10
4.5 Chapter 5 of the Thesis: Consensus Meeting	10
4.6 Chapter 6 of the Thesis: Draft Checklist and E&E Document	11
4.7 Chapter 7 of the Thesis: Dissemination & Implementation Plan	11
4.8 Chapter 8 of the Thesis: Conclusions about the CONSORT-SPI Project	12
References	13
Publications and Presentations Related to CONSORT-SPI to Date	18

<b><i>Chapter 1: Randomised Controlled Trials of Social and Psychological Interventions</i></b>	
1. Introduction	20
2. Scientific Approaches to Establishing Causal Relationships	20
2.1 Scientific Experiments	21
2.2 Randomised Clinical Trials in Medicine	22
2.3 Evidence-Based Medicine	24
2.4 The Evidence Hierarchy of EBM	25
<b>Figure 1. EBM Hierarchy of Evidence for Determining Intervention Effectiveness</b>	<b>26</b>
3. The Evidence-Based Practice Movement	26
3.1 Experimentation in the Social and Behavioural Sciences	26
3.2 Emergence of Evidence-Based Practice	27
3.3 EBP for Social and Psychological Interventions	27
4. Evaluating Social and Psychological Interventions with RCTs	29
4.1 Characteristics of Social and Psychological Interventions	30
<b>Box 1: Features of Social and Psychological Interventions</b>	<b>31</b>
4.1.1 Intervention Theory and Mechanisms of Action	31
4.1.2 Intervention Implementation	32
4.1.3 Contextual Dependence of Intervention Effects	32
4.2 Design Features for RCTs of Social and Psychological Interventions	33
<b>Figure 2. Key Elements of the Updated MRC Framework for the Development and Evaluation of Complex Interventions</b>	<b>34</b>
4.2.1 RCT Design Features Related to Internal Validity	35
4.2.2 RCT Design Features Related to External Validity	37
5. Reporting Standards and Guidelines for Research	40
5.1 The Need for Transparent Reporting	40
5.2 Evidence of Poor Reporting	42
5.3 Reporting Guidelines for Health Research	44
<b>Figure 3: CONSORT 2010 Statement Checklist</b>	<b>46</b>
<b>Figure 4: CONSORT 2010 Statement Flowchart</b>	<b>48</b>
5.4 A CONSORT Extension for Social and Psychological Intervention RCTs?	49
<b>Table 1. Social and Psychological Interventions from the Campbell Library</b>	<b>51</b>
6. Conclusion	53
References	55

## ***Chapter 2: Protocol for Developing the CONSORT-SPI Extension***

1. Introduction	67
1.1 Objective	67
1.2 Project Overview and Thesis Outline	68
<b>Figure 1. Workflow for Thesis on CONSORT-SPI Project</b>	<b>69</b>
1.3 Feasibility	71
1.3.1 Project Executive	71
1.3.2 International Advisory Group	71
1.3.3 Journal Editors	72
1.3.4 CONSORT Group	72
1.4 DPhil Activities in the CONSORT-SPI Project	73
2. Project Phase 1: Systematic Reviews	73
2.1 Objective	73
2.2 Thesis Chapter 3: Protocol for Systematic Reviews	73
2.2.1 Review of Relevant Reporting Guidelines	74
2.2.2 Review of RCT Reporting Quality	74
3. Project Phase 2: The Delphi Process	75
3.1 Objective	75
3.2 Thesis Chapter 4: The CONSORT-SPI Delphi Process Protocol	76
3.2.1 International Advisory Group Meeting	77
3.2.2 Selection of Participants	77
3.2.3 Structure	78
4. Project Phase 3: Consensus Development Conference	81
4.1 Objective	81
4.2 Thesis Chapter 5: Protocol for Consensus Meeting Methods	82
4.2.1 International Advisory Group Meeting	82
4.2.2 Selection of Participants	82
4.2.3 Structure	82
5. Project Phase 4: Drafting the Guideline Documents	84
5.1 Thesis Chapter 6: Protocol for CONSORT-SPI E&E Template Development	84
5.1.1 Draft Explanation and Elaboration Document	84
6. Project Phase 5: Guideline Dissemination	85
6.1 Thesis Chapter 7: Protocol for the Development of the Dissemination Strategy	85

6.1.1 Insights from Previous Reporting Guidelines and the CONSORT-SPI Consensus Meeting	85
6.1.2 Review of Implementation Science for Clinical Guidelines	86
7. Thesis Chapter 8: Conclusion	86
References	88

### *Chapter 3:*

#### *Reporting Quality of Social and Psychological Intervention Trials: A Systematic Review of Reporting Guidelines and Trial Publications*

Abstract	93
1. Introduction	95
2. Methods	97
2.1 Eligibility Criteria	97
2.2 Search Strategy and Study Selection	98
<b>Box 1. Electronic Search Strategy Adapted from Moher 2011</b>	<b>99</b>
2.3 Data Extraction	103
2.4 Data Analysis	105
3. Results	106
3.1 Previous Guidance for Social and Psychological Intervention Trials	106
<b>Figure 1. Flowchart of Records through Literature Search for Reporting     Guidance</b>	<b>107</b>
<b>Table 1. Characteristics of Included Reporting Guidelines and     Quality Assessment Tools</b>	<b>109</b>
<b>Table 2. Average Percentage of Employed Recommended     Practices for Guideline Development by Document Type</b>	<b>110</b>
3.2 Assessment of Reporting Quality of Social and Psychological Intervention Trials	111
3.2.1 Reporting Characteristics: General Reporting Quality	111
<b>Figure 2. Flowchart of Considered Publications through Review     of Reporting Quality</b>	<b>112</b>
3.2.2 Reporting Characteristics: Title, Abstract, and Introduction	113
3.2.3 Reporting Characteristics: Methods	113
3.2.4 Reporting Characteristics: Results	114
3.2.5 Reporting Characteristics: Discussion	114
<b>Table 3. Sample of Journals Included in Reporting Quality Review</b>	<b>115</b>
<b>Table 4. Average Compliance of 309 RCTs with Reporting Standards</b>	<b>118</b>
3.2.6 Reporting Characteristics: Important Study Details	121
4. Discussion	121
4.1 Overall Findings	121
4.2 Guideline Adherence to Best-Practices	122

4.2.1 Best-Practices for Preliminary Planning	123
4.2.2 Best-Practices for Consensus Development	124
4.2.3 Best-Practices for Publication and Dissemination	126
4.3 Reporting Standards for Social and Psychological Intervention Trials	127
4.3.1 Internal Validity	127
4.3.2 External Validity	128
4.3.3 Study Details	130
4.4 Strengths and Limitations of the Current Study	131
4.5 Implications	133
5. Conclusion	135
References	136

**Chapter 4:**  
***Identifying Reporting Standards for Social and Psychological Intervention Trials:  
An Online, International Delphi Process***

Abstract	142
1. Introduction	144
2. Methods	146
2.1 Selection of Participants	147
2.2 Selection of Preliminary Items	149
2.3 Round 1 Delphi Survey	150
2.3.1 Procedures	150
2.3.2 Materials	151
2.3.3 Data Analysis	152
2.4 Round 2 Delphi Survey	154
2.4.1 Procedures	154
2.4.2 Materials	154
2.4.3 Data Analysis	156
3. Results	158
3.1 Delphi Participants	158
<b>Figure 1. CONSORT-SPI Delphi Participant Flow Diagram</b>	<b>159</b>
<b>Table 1. Gender, Age, Country, and Discipline of Delphi Participants</b>	<b>160</b>
<b>Table 2. Professional Roles of Delphi Participants</b>	<b>161</b>
<b>Table 3. Number of Intervention-Related Publications by                 Academics/Researchers in Delphi Process</b>	<b>162</b>
<b>Table 4. Years of Experience by Professionals in Delphi Process</b>	<b>163</b>
3.2 Round 1 Results	164
3.2.1 Quantitative Rankings	164
3.2.2 “Title and Abstract” Text Responses	164
3.2.3 “Introduction” Text Responses	164
3.2.4 “Trial Design” Text Responses	165
3.2.5 “Randomisation, and Blinding Procedures” Text Responses	165
3.2.6 “Methods: Process Evaluation” Text Responses	166
3.2.7 “Methods: Outcomes and Data Analysis” Text Responses	168
3.2.8 “Results: Participant Flow and Recruitment” Text Responses	168
3.2.9 “Results: Process Evaluation” Text Responses	169

3.2.10 “Results: Outcomes and Estimation” Text Responses	170
3.2.11 “Discussion” Text Responses	170
3.2.12 “Other Information” Text Responses	171
3.2.13 Additional Items	172
<b>Table 5. Items Recommended for Inclusion in the CONSORT-SPI Checklist from Round 1</b>	<b>173</b>
<b>Table 6. Indeterminate Items from Round 1</b>	<b>175</b>
3.3 Round 2 Results	178
3.3.1 Quantitative Rankings	178
3.3.2 Text Responses for Items Recommended for Inclusion in Round 1	178
3.3.3 “Title and Abstract” Text Responses	179
3.3.4 “Introduction” Text Responses	180
3.3.5 “Trial Design, Randomisation, and Blinding Procedures” Text Responses	181
3.3.6 “Trial Setting” Text Responses	181
3.3.7 “Interventions” Text Responses	182
3.3.8 “Outcomes and Data Analysis” Text Responses	183
3.3.9 “Baseline and Outcomes Data” Text Responses	184
3.3.10 “Process Evaluation” Text Responses	185
3.3.11 “Discussion” Text Responses	186
3.3.12 “Other Information” Text Responses	186
3.3.13 Text Responses for New Items Suggested in Round 1	187
3.3.14 Superordinate Themes from Delphi Process Text Responses	188
4. Discussion	188
4.1 Overall Findings	188
<b>Table 7. Recommended Items for the CONSORT-SPI Checklist from the Delphi Process</b>	<b>189</b>
<b>Table 8. Indeterminate Items from the Delphi Process</b>	<b>193</b>
4.2 Strengths and Limitations of the Current Study	196
4.3 Implications and Future Directions	198
References	200

**Chapter 5:**  
***The CONSORT-SPI Consensus Meeting:***  
***Developing the Guideline Checklist***

Abstract	204
1. Introduction	206
2. Methods	207
2.1 Eligibility Criteria	207
2.2 Recruitment Strategy	207
2.3 Time and Setting	208
2.4 Meeting Preparation	208
2.5 Structure of Meeting	209
2.5.1 Day 1 Agenda	210
2.5.2 Day 2 Agenda	211
2.5.3 Day 3 Agenda	212
2.6 Data Analysis	213
3. Results	215
3.1 Flow of Participants and Items	215
3.2 Day 1 Meeting Minutes	215
3.2.1 Session 1.1—Welcome & Introductions	215
<b>Figure 1. CONSORT-SPI Consensus Meeting Participant Flow Diagram</b>	<b>216</b>
<b>Table 1. Votes on Proposed CONSORT-SPI Checklist Items</b>	<b>217</b>
<b>during the CONSORT-SPI Consensus Meeting</b>	
3.2.2 Session 1.2—Overview of Goals and Process	220
3.2.3 Session 1.3—Background: Reporting Guidelines	222
3.2.4 Session 1.4—Results from Systematic Reviews	226
3.2.5 Session 1.5— Group Discussion: Reporting Problems in “Your” Area	229
3.2.6 Session 2.1—Focus of Checklist	234
3.2.7 Session 2.2—The CONSORT-SPI Delphi Process	236
3.2.8 Session 2.3—Discussion: Results from the Delphi Process	237
3.3 Day 2 Meeting Minutes	239
3.3.1 Session 3.1—Introduction Section	239
3.3.2 Session 3.2—Trial Design	243
3.3.3 Session 3.3—Randomisation Procedures	245
3.3.4 Session 3.4—Blinding Procedures	246

3.3.5 Session 3.5—Intervention Design	248
3.3.6 Session 3.6—Intervention Implementation: Setting, Delivery, and Uptake	251
3.3.7 Session 4.1—Data Collection and Analytic Plan	253
3.3.8 Session 4.2—Participant Flow and Recruitment	256
3.3.9 Session 4.3—Baseline Data	259
3.3.10 Session 4.4—Results: Outcomes and Estimation	261
3.3.11 Session 4.5—Adverse Events	263
3.3.12 Session 4.6—Discussion Section	263
3.3.13 Session 4.7—“Other” Information	265
3.4 Day 3 Meeting Minutes	268
3.4.1 Session 5.1— Review of Day 2 and Aims of Day 3	268
3.4.2 Session 5.2—Tabled Issues from Day 2	269
3.4.3 Session 5.3—Review and Discussion of Recommended Items	269
3.4.4 Session 5.4— Finalise CONSORT-SPI Checklist	274
3.4.5 Session 5.5— Items for Title and Abstract	276
3.4.6 Session 5.6— Discussion on Checklist Wording	277
4. Discussion	278
4.1 Overall Findings	278
4.2 Strengths and Limitations of the Current Study	279
4.3 Implications and Future Directions	281
References	282

**Chapter 6:**  
***CONSORT-SPI Explanation and Elaboration:***  
***Template for the Checklist Users Manual***

Abstract	286
1. Introduction	288
2. Methods	289
2.1 Procedure	289
2.2 Preambles to the E&E Requested by Consensus Meeting Participants	290
2.2.1 Remit of CONSORT-SPI	290
<b>Table 1. Social and Psychological Interventions from the Campbell Library</b>	<b>291</b>
2.2.2 Other CONSORT Extensions	292
2.2.3 Nature of the Intervention Techniques and the Trial	293
2.2.4 Meaning of “Participants” in the CONSORT-SPI Checklist	293
2.2.5 Intervention Implementation and Process Evaluations	294
2.2.6 Record Linkage and Sibling Documents	295
2.3 Changes to the CONSORT-SPI Checklist After the Consensus Meeting	296
<b>Table 2. Draft CONSORT-SPI Checklist (Stage: Post-Consensus Meeting)</b>	<b>297</b>
3. Results: CONSORT-SPI Explanation and Elaboration	302
3.1 Title and Abstract	302
3.2 Introduction	304
<b>Box 1. CONSORT for Abstracts Checklist</b>	<b>305</b>
3.3 Methods: Trial Design	309
<b>Figure 1. Example Logic Model for a Complex Social Intervention</b>	<b>310</b>
<b>Table 3. Extended Items from CONSORT for Cluster Randomised Trials</b>	<b>312</b>
3.4 Methods: Participants	313
3.5 Methods: Interventions	316
<b>Table 4. Template for Intervention Description and Replication (TIDieR)</b>	<b>318</b>
3.6 Methods: Outcomes	325
3.7 Methods: Sample Size	328
3.8 Methods: Randomisation—Sequence Generation	330
3.9 Methods: Randomisation—Allocation Concealment Mechanism	331
3.10 Methods: Randomisation—Implementation	332
3.11 Methods: Awareness of Assignment	332
3.12 Methods: Analytical Methods	335

3.13 Results: Participant Flow	338
<b>Figure 2. Example Flow Diagram From Number Approached to Analysed for Primary Outcome</b>	<b>340</b>
<b>Figure 3. Example Flow Diagram with Information on Participants Discontinuing Intervention but Remaining in Analysis</b>	<b>341</b>
<b>Figure 4. Example Flow Diagram for Cluster SPI Trials</b>	<b>342</b>
3.14 Results: Recruitment	343
3.15 Results: Baseline Data	344
<b>Table 5. Example Table Showing Baseline Demographics</b>	<b>345</b>
3.16 Results: Numbers Analysed	346
<b>Table 6. Example Table with Number of Participants Per Group</b>	<b>347</b>
3.17 Results: Outcomes and Estimation	348
<b>Table 7. Example of Reporting of Summary Results for Each Study Group</b>	<b>349</b>
<b>Table 8. Example of Summary Results per Group Helpful for Meta-Analysts</b>	<b>350</b>
<b>Table 9. Example of Reporting Both Relative Effect Sizes and Data for Calculating Absolute Effects</b>	<b>353</b>
3.18 Results: Ancillary Analyses	354
3.19 Results: Harms	354
3.20 Discussion: Limitations	355
3.21 Discussion: Generalisability	356
3.22 Discussion: Interpretation	358
3.23 Important Information: Registration	359
3.24 Important Information: Protocol	360
3.25 Important Information: Funding	361
3.26 Important Information: Stakeholder Involvement	362
4. Discussion	365
4.1 Strengths and Limitations	365
4.2 Implications	366
5. Conclusion	367
References	368

## *Chapter 7:*

### *A Conceptual Framework for Disseminating and Implementing Reporting*

#### *Guidelines: An Example using CONSORT-SPI*

Abstract	379
1. Introduction	381
2. Methods	382
2.1 Materials: Guidance for Disseminating Reporting Guidelines	383
2.2 Materials: Original CONSORT-SPI Dissemination and Implementation Strategy	385
<b>Box 1. Initial Pathways to Impact Plan for CONSORT-SPI</b>	<b>386</b>
2.3 Materials: CONSORT-SPI Consensus Meeting Discussion	388
2.4 Procedure	390
2.4.1 Step 1. Who Needs to Do What, Differently?	390
2.4.2. Step 2. Using a Theoretical Framework, which Barriers and Enablers Need to be Addressed?	390
2.4.3 Step 3. Which Intervention Components Could Overcome the Modifiable Barriers and Enhance the Enablers?	393
2.4.4 Step 4. How Can Behaviour Change be Measured and Understood?	393
3. Results	394
3.1 Step 1. Who Needs to Do What, Differently?	394
3.2 Step 2. Using a Theoretical Framework, which Barriers and Enablers Need to be Addressed?	396
3.2.1 Barriers and Enablers of the Guideline Itself	396
<b>Table 1. Potential Barriers and Enablers of the Guideline Itself</b>	<b>397</b>
3.2.2 Individual-Level Barriers and Enablers	398
<b>Table 2. Potential Individual-Level Barriers and Enablers before         Using CONSORT-SPI</b>	<b>399</b>
<b>Table 3. Potential Within-Individual Barriers and Enablers while         Using CONSORT-SPI</b>	<b>400</b>
<b>Table 4. Potential Contextual Barriers and Enablers during         Individuals' Use of CONSORT-SPI</b>	<b>402</b>
3.2.3 Organisational-Level Barriers and Enablers	403
<b>Table 5. Potential Organisational-Level Barriers and Enablers before         Implementing CONSORT-SPI</b>	<b>404</b>

<b>Table 6. Potential Within-Organisation Barriers and Enablers while Using CONSORT-SPI</b>	<b>406</b>
3.2.4 System-Level Barriers and Enablers	407
<b>Table 7. Potential Contextual Barriers and Enablers during Organisational Implementation of CONSORT-SPI</b>	<b>408</b>
<b>Table 8. Potential Community-Level Barriers and Enablers before Supporting Reporting Guidelines</b>	<b>409</b>
<b>Table 9. Potential Within-Community Barriers and Enablers while Supporting Reporting Guidelines</b>	<b>411</b>
<b>Table 10. Potential Contextual Barriers and Enablers during Community Support of Reporting Guidelines</b>	<b>413</b>
3.3 Step 3. Which Intervention Components Could Overcome the Modifiable Barriers and Enhance the Enablers?	414
3.3.1 Guideline-Level Implementation Intervention Components	414
3.3.2 Individual-Level Implementation Intervention Components	416
3.3.3 Organisational-Level Implementation Intervention Components	417
3.3.4 System-Level Implementation Intervention Components	419
4.4 Step 4. How Can Behaviour Change Be Measured and Understood?	420
4. Discussion	421
4.1 Strengths and Limitations	422
4.2 Conclusions	423
References	425

## ***Chapter 8:***

### ***Conclusions about the CONSORT-SPI Project***

Overview	430
1. What We Already Knew	431
2. What This Thesis Adds	432
2.1 Systematic Reviews	432
2.2 Delphi Process	433
2.3 Consensus Meeting	434
2.4 E&E Document	435
2.5 Dissemination and Implementation Strategy	435
3. Strengths and Limitations of this Thesis	436
4. Implications for Future Research	441
4.1 Future Research for Social and Psychological Intervention Researchers	442
4.2 Other Reporting Guidelines	444
5. Closing	445
References	447



## *Preface to DPhil Thesis*

### **1. What are Social and Psychological Interventions?**

Broadly defined, an intervention is “the action of intervening, ‘stepping in’, or interfering in any affair, so as to affect its course or issue.”<sup>1</sup> This broad definition can help with operationalising the concepts that are often used to distinguish interventions in the medical, social, and behavioural sciences. Namely, these interventions can be thought to involve techniques (the “actions”) that aim to modify an otherwise unhindered process or phenomenon (the “affair”) so as to improve the functioning and well-being of the unit of interest (the “course or issue”). Interventions can be logically classified in a variety of ways across these domains. For example, current intervention extensions of the Consolidated Standards of Reporting Trials (CONSORT) Statement are organised by intervention techniques: compared to interventions involving pharmacologic techniques or drugs,<sup>2</sup> other interventions are differentiated by their use of herbal medicinal products,<sup>3</sup> acupuncture,<sup>4</sup> or non-pharmacologic techniques more generally.<sup>5</sup>

Rather than classifying interventions according to the techniques employed, this thesis distinguishes the interventions of interest by the mechanisms targeted: i.e., the processes and phenomena that interventionists intend to modify in order to affect certain outcomes. Such classification corresponds to current “intervention development” approaches that promote the identification of malleable risk and protective factors as a precursor to postulating which techniques may best address a problem or outcome of interest.<sup>6,7</sup> Using this classification system, “biomedical” interventions are those that target and are conceived to work through biological mechanisms, such as physiological processes. In comparison, “social and psychological” interventions target and are conceived to work through malleable mental processes and social phenomena, such as cognitions, emotions, behaviours, norms, interpersonal relationships, and salient physical

and social environments, amongst others. The distinctive concepts, theories, and taxonomies characteristic of social and psychological intervention research are central to the rationale for this thesis project.

The DPhil candidate acknowledges that, in reality, there are not strict boundaries amongst the “biological”, the “psychological”, and the “social”, but rather all are intertwined in some fashion. Nonetheless, in practice there are real differences in the conceptualisation and nature of the interventions that are developed and evaluated by those with predominantly biomedical backgrounds compared to those whose principal training is in the social and behavioural sciences. Though a non-exhaustive list, social and psychological interventions are studied in several disciplines, including criminology, education, psychology, public health, and social work. Across these disciplines, social and psychological interventions distinguish themselves from biomedical interventions in several ways.

For example, conceiving of and focusing on social and psychological mechanisms lead to quite different techniques than those found in biomedical interventions. The area of social and psychological intervention research involves a distinct assortment of activities—psychotherapies, programmes, services, projects, policies—that target mental processes and social phenomena across individuals, groups, whole populations, and even places. The targeted mechanisms are inseparable from the agency of those actively receiving or engaging with the intervention. As these participants may differentially respond to intervention techniques or activities, researchers must take account of human agency and intentionality in their theory of how social and psychological interventions work. Consequently, stakeholders in social and psychological intervention research are interested in more than just overall effect estimates of interventions, seeking further information on how and why these interventions work, for whom, and under what conditions.

## 2. Randomised Controlled Trials of Social and Psychological Interventions

Social and psychological interventions are designed to improve a variety of outcomes for people and communities. Though developed and provided with the best of intentions, these interventions have the potential to be ineffective or inadvertently harmful.<sup>8</sup> Social and psychological interventions therefore need to demonstrate evidence of safety and effectiveness as a prerequisite for wide-scale delivery.<sup>9</sup>

Several methods exist to provide such evidence. When appropriately used and conducted, randomised controlled trials (RCTs) are considered by many to be the most valid research method for estimating the effectiveness of interventions. Selection biases can lead to systematic differences in prognostic variables between participants in different experimental groups,<sup>10</sup> and random assignment prevents biases related to the selection of participants for intervention and comparator groups. Because RCTs theoretically balance both measured and unmeasured prognostic variables, they can provide valid measurements of the impact of an intervention compared to what would have happened in its absence.<sup>11</sup>

Over the past several decades, RCTs have been used to evaluate ever more complex interventions, such as social and psychological interventions, that challenge traditional aspects of trial design, execution, and interpretation.<sup>7,12</sup> Experiences of evaluating complex interventions have led to an accumulation of innovative modifications to conventional RCT methodology. While debate has since arisen (and continues today) about the appropriateness of RCTs for evaluating these interventions,<sup>13</sup> solutions to the technical and ethical problems of such trials are achieving increasing consensus. Agreement on these solutions has reached the point that influential guidance has been developed for those conducting evaluations of complex interventions.<sup>7,14,15</sup> For instance, RCTs of social and psychological interventions are increasingly interpreted to test whether the *offer* of an intervention makes a difference, and, via an embedded process evaluation, they may

investigate sources of variation in response to the intervention on offer.<sup>16,17</sup> Though discussed in more detail in Chapter 1, it is worth noting now that this thesis does not aim to make a case for a hierarchy of evidence that always privileges on first principles the use of an RCT to evaluate social and psychological interventions over other methods. Rather, the author acknowledges the RCT as an essential tool in the evaluation toolkit, and that how RCTs are actually done may benefit from monitoring and oversight given their prominence in the evidence-based practice (EBP) paradigm at the time of the project.

### **3. Reporting Standards and Reporting Guidelines**

Growing attention to the *design* and *conduct* of RCTs of complex interventions has led to growing attention to the *reporting* of these studies in scientific journals and other media. Reports are the primary means of distributing RCT findings to the research community.<sup>18</sup> Research consumers depend on accurate, complete, and transparent reports of RCTs in order to appraise the validity, applicability, and appropriateness of individual studies.

To address the need for high quality reporting, collaborations led by researchers and journal editors have made a concerted effort over the last 20 years to develop reporting standards that assist authors in writing research manuscripts. Reporting standards provide recommendations to study authors for describing the most important aspects of empirical research studies. They are typically collated and disseminated through reporting guidelines: i.e., checklists that offer recommendations on what to report about a particular type of study.<sup>19</sup> Ideally, standards found in reporting guidelines should be based on previous research rather than guideline developers' intuition alone, and they should also be developed using expert consensus to improve buy-in amongst key stakeholders and to reduce biases in group decision-making.<sup>20</sup>

### ***3.1 The CONSORT Statement***

The most well-known and widely used reporting guideline is the Consolidated Standards of Reporting Trials (CONSORT) Statement. The CONSORT 2010 Statement—the most recent version—provides a set of recommendations primarily for reports of two-group, parallel RCTs.<sup>2</sup> To increase the usefulness of this guideline, the CONSORT Group has also produced extended versions for specific types of medical interventions (such as non-pharmacological treatments<sup>21</sup>), for specific types of RCTs (such as cluster trials<sup>22</sup>), and for specific types of RCT data (such as harms<sup>23</sup>). The reporting standards found within the CONSORT Statement and its official Extensions are based on empirical data and expert consensus about the potential sources of bias in RCTs.<sup>2</sup> Over the last decade, numerous reviews in the biomedical literature have attributed improvements in the reporting quality of RCTs to these guidelines.<sup>24</sup>

### ***3.2 A CONSORT Extension for Social and Psychological Interventions?***

Despite the advance of CONSORT and other reporting guidelines, several studies indicate that there are persistent deficiencies in the reporting quality of published social and psychological intervention RCTs.<sup>25-29</sup> One explanation may be that current standards in well-known reporting guidelines are not adequately tailored to social and psychological intervention trials. For example, existing CONSORT guidance has been developed primarily by biomedical researchers, making guidance for social and behavioural science research lacking in comparison. The main CONSORT 2010 Statement was designed with two-group parallel RCTs of pharmaceuticals in mind,<sup>2</sup> and existing intervention extensions of CONSORT are intended for non-pharmacologic medical treatments (e.g., surgery, medical devices),<sup>5</sup> herbal medicinal products,<sup>3</sup> and acupuncture treatments.<sup>4</sup> All of these interventions primarily target mechanisms emphasised in various areas of medical science,

both traditional Western medicine as well as the “alternative and complementary”.

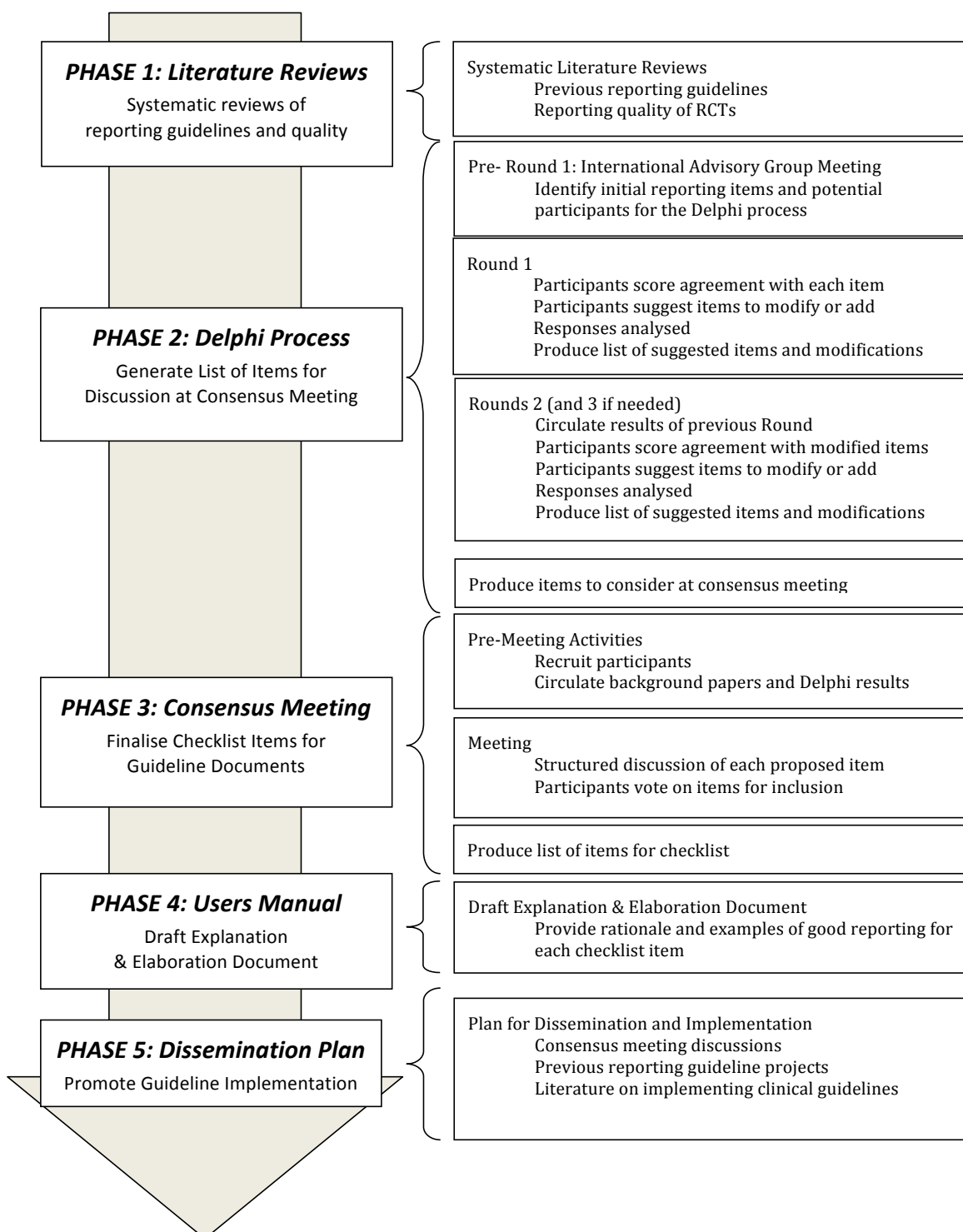
Interventions targeting mechanisms common in the social and behavioural scientists are underrepresented in the CONSORT family of guidelines. Many important concepts, theories, and taxonomies characteristic of social and psychological intervention research may consequently be overlooked by these guidelines to some degree. Indeed, stakeholders in these areas have indicated that current CONSORT guidelines do not appear to sufficiently target social and psychological intervention RCTs.<sup>28,30,31</sup> Members of previous CONSORT groups, as well as journal editors and researchers in the social and behavioural sciences, support the inclusion of stakeholders in social and psychological intervention research in future guideline development to improve “buy-in” and relevance of CONSORT to these disciplines moving forward.<sup>28,32</sup> The need for a new CONSORT extension for these trials should therefore be explored.

#### **4. Overview of Thesis Structure**

This DPhil thesis reports on the development of an official CONSORT Extension for Social and Psychological Interventions: CONSORT-SPI. The project plan for developing this CONSORT extension involves five phases (see Figure 1).

As requested during the DPhil candidate’s confirmation of status viva examination, it is worth noting here the collaborative nature of this project and the DPhil candidate’s roles within it. The idea for the project was conceived by Paul Montgomery and Evan Mayo-Wilson prior to the DPhil candidate’s involvement. After being approached by Prof Montgomery and Dr Mayo-Wilson to explore this idea further, the DPhil candidate has been involved in a lead role at every phase of the project. Namely, the DPhil candidate helped to develop the conceptual rationale for the project as well as the grant proposals to the UK Medical Research Council (unsuccessful) and Economic and Social Research

**Figure 1. Workflow for Thesis on CONSORT-SPI Project**



Council (successful). The DPhil candidate wrote the initial draft for both the conceptual rationale<sup>33</sup> as well as the grant proposals, which were then refined by the project team. The DPhil candidate also wrote the first draft of the project protocol for his transfer of status examination; this protocol was subsequently revised for publication.<sup>34</sup> The content of the subsequent thesis chapters depicts the main role of the DPhil candidate in leading the design and execution of the research methods constituting each phase of the project. The specific activities of the DPhil candidate will be noted throughout the thesis as relevant, and a list of publications and presentations to date has been included in the reference section of this preface.

Several appendices have been included in this thesis submission, also at the behest of the confirmation of status viva examiners. These appendices provide some of the materials developed by the DPhil candidate for the project, full data from certain project phases, and the evolution of the draft CONSORT-SPI checklist. The confirmation of status examiners agreed that, in accord with the principles of transparency underpinning this project, there should be open access to data and materials from this thesis.

#### ***4.1 Chapter 1 of the Thesis: Conceptual Overview***

NB: Sections of this chapter have been published in an alterative form in *Research on Social Work Practice*<sup>33</sup>

The first chapter is an overview of the developments in intervention research that led to innovative uses of the RCT to evaluate social and psychological interventions. Namely, the DPhil candidate highlights how RCTs evaluating more complex interventions have sought to combine the advantages of random allocation for outcome evaluations with much greater attention to process and contextual aspects of the evaluation in order to examine how and why the intervention works, for whom, and under what conditions.<sup>13,35</sup> The DPhil

candidate then argues that researchers may require updated and tailored reporting guidelines for social and psychological intervention trials if these studies are to be useful for policy, practice, and future research. The purpose of this chapter is to situate this DPhil project in the wider scientific literature. The chapter highlights the conceptual arguments indicating that the need for a new reporting guideline for social and psychological intervention RCTs should be formally investigated.

#### ***4.2 Chapter 2 of the Thesis: Project Protocol***

NB: This chapter has been published in an alternative form in *Implementation Science*<sup>34</sup>

The second chapter delineates the methods for the overall project and the corresponding activities within the project plan that will constitute chapters for this thesis. The proposed methods follow recommended techniques for guideline development and dissemination.<sup>19</sup> The project received a standard grant from the Economic and Social Research Council (ESRC) and ethics approval from the Departmental Research Ethics Committee (DREC), Department of Social Policy and Intervention, University of Oxford, for the proposed methods. Moreover, to help coordinate and publicise the initiative, an International Advisory Group (IAG) of leading intervention experts in core disciplines was assembled (see Appendix A).

#### ***4.3 Chapter 3 of the Thesis: Systematic Reviews***

NB: This chapter has been published in an alternative form in *PLoS One*<sup>36</sup>

This chapter reports on Phase 1 of the project: a study consisting of two systematic reviews that sought to empirically demonstrate the need for a CONSORT extension for social and psychological interventions. Overall, these reviews examined reporting guidelines for, and the current reporting quality of, social and psychological intervention

RCTs. Firstly, a systematic review identified and assessed the rigour of all published reporting guidelines relevant to social and psychological intervention RCTs. The quality of guideline development, content (by way of included reporting standards), and dissemination strategies were assessed according to recommended techniques for guideline production.<sup>19</sup> A second review examined the extent to which roughly 300 social and psychological intervention RCT publications adhered to the reporting standards that were identified in the review of reporting guidance. Results from this study suggest that current reporting guidelines are insufficient for social and psychological intervention trials, and a new CONSORT extension would likely improve their reporting quality.

#### ***4.4 Chapter 4 of the Thesis: Delphi Process***

This chapter reports on Phase 2 of the project: an international, online Delphi process to generate a prioritised list of possible reporting standards to consider for inclusion in the CONSORT extension. Information from the two-part study in Phase 1 informed the content of the Round 1 questionnaire of the Delphi process. Participants consisted of researchers, journal editors, research funders, policy-makers, practitioners, and other stakeholders in social and psychological interventions. Results from the Delphi process led to a ranked, modified list of reporting items that were considered for inclusion in the CONSORT-SPI checklist and E&E document at the consensus meeting.

#### ***4.5 Chapter 5 of the Thesis: Consensus Meeting***

This chapter reports on Phase 3 of the project: a formal consensus meeting to agree upon the reporting items to add to or modify for the CONSORT-SPI checklist. A group of 31 researchers, journal editors, and funders met in March 2014 during a three-day consensus development conference. This meeting involved discussion of preliminary

research on social and psychological intervention RCTs, voting on items for the CONSORT-SPI checklist, and deliberations on the dissemination and implementation plan for the CONSORT-SPI guidelines. The CONSORT-SPI checklist is currently planned to consist of 14 extended items of the CONSORT 2010 checklist, as well as a flow diagram for tracking participants through an RCT. This chapter reports on the discussions and decisions made during the meeting relating to the CONSORT-SPI checklist. Though related to the consensus meeting, a draft of the CONSORT-SPI checklist and E&E document will be in Chapter 6 of the thesis, and an overview of discussions at the consensus meeting about the dissemination and implementation of CONSORT-SPI will be in Chapter 7.

#### ***4.6 Chapter 6 of the Thesis: Draft Checklist and E&E Document***

This chapter relates to Phase 4 of the project and includes a template for the CONSORT-SPI E&E document, according to the most recent draft of the CONSORT-SPI checklist. This phase involved revising and codifying the standards finalised at the consensus meeting into a user-friendly checklist, and writing an explanation and elaboration (E&E) document that provides detailed advice and examples of good reporting for each reporting standard. A detailed examination of each of the reporting standards within the new CONSORT extension constitutes this chapter of the thesis.

#### ***4.7 Chapter 7 of the Thesis: Dissemination & Implementation Plan***

This thesis chapter reports on the fifth and final phase of the project: the guideline dissemination and implementation plan. The DPhil candidate will discuss a framework for a coordinated dissemination and implementation strategy that is based on discussions at the CONSORT-SPI consensus meeting, previous plans for reporting guidelines, and insights

derived from literature on implementing clinical guidelines and empirically supported interventions. This plan includes simultaneous publication of the guideline in multiple journals, guideline endorsement by journals and funding organisations, presentations at conferences and organisational meetings, and a dedicated website allowing for feedback about the guideline. Other policies supportive of high-quality scientific publishing more generally are also discussed.

#### ***4.8 Chapter 8 of the Thesis: Conclusions about the CONSORT-SPI Project***

The final chapter of the thesis details conclusions about the CONSORT-SPI project to date. The DPhil candidate discusses what was previously known about the topic prior to the project; what this thesis adds to the literature; strengths and limitations of this thesis project; and implications of the thesis work for policy, practice, and future research.

## References

1. intervention, n. *OED Online*. Oxford: Oxford University Press; March 2014.
2. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698-702.
3. Gagnier J, Boon H, Rochon P, Barnes J, Moher D, Bombardier C. Reporting random controlled trials of herbal medicines. *Explore: The Journal of Science and Healing*. 2006;2(2):143-149.
4. MacPherson H, White A, Cummings M, et al. Standards for Reporting Interventions in Controlled Trials of Acupuncture: The STRICTA recommendations. *Journal of Alternative & Complementary Medicine*. 2002;8(1):85-89.
5. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT Group. Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: Explanation and Elaboration. *Annals of Internal Medicine*. 2008;148:295-309.
6. Fraser MW, Galinsky MJ, Richman JM, Day SH. *Intervention research: Developing social programs*. New York, NY: Oxford University Press; 2009.
7. Medical Research Council. *Developing and evaluating complex interventions: New guidance*. London: Medical Research Council; 2008.
8. McCord J. Cures that harm: Unanticipated outcomes of crime prevention programs. *The ANNALS of the American Academy of Political and Social Science*. 2003;587:16-30.

9. Flay BR, Biglan A, Boruch RF, et al. Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*. 2005;6:151-175.
10. Jun M, Zoungas S, Perkovic V, Webster AC. How to read a report of a randomized controlled trial. *Nephrology*. 2010;15:153-157.
11. Chalmers I. Trying to do more good than harm in policy and practice: The role of rigorous, transparent, up-to-date evaluations. *The ANNALS of the American Academy of Political and Social Science*. 2003;589(22):22-40.
12. Shepperd S, Lewin S, Straus S, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Medicine*. 2009;6(8):31000086.
13. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: Suggested framework and systematic review. *BMJ*. 2006;333:346-349.
14. Goldacre B. *Bad science*. London: Harper Perennial; 2009.
15. Torgerson DJ, Torgerson CJ. *Designing randomised trials in health, education and the social sciences: An introduction*. Palgrave Macmillan; 2008.
16. Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*. 2012;75(12):2299-2306.
17. Weiss MJ, Bloom HS, Brock T. *A conceptual framework for studying the sources of variation in program effects*. New York, NY: MDRC; 2013.
18. Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR Network's Survey of Guideline Authors. *PLoS Medicine*. 2008;5:869-874.
19. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.

20. Bennett C, Khangura S, Brehaut JC, et al. Reporting guidelines for survey research: An analysis of published guidance and reporting practices. *PLoS Medicine*. 2011;8(8):e1001069.
21. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
22. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: Extension to cluster randomised trials. *BMJ*. 2004;328:702-708.
23. Ioannidis JPA, Evans SJW, Gotszche PC, et al. Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Ann Intern Med*. 2004;141:781-788.
24. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *MJA*. 2006;185:263-267.
25. Perry AE, Weisburd D, Hewitt C. Are criminologists describing randomized controlled trials in ways that allow us to assess them? Findings from a sample of crime and justice trials. *Journal of Experimental Criminology*. 2010;6:245-262.
26. Han C, Kwak K, Marks DM, et al. The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. *Contemporary Clinical Trials*. 2009;30:116-122.
27. Naleppa MJ, Cagle JG. Treatment fidelity in social work intervention research: A review of published studies. *Research on Social Work Practice*. 2010;20:674-681.

28. Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. A comparison of RCTs in health and education. *British Educational Research Journal*. 2005;31(6):761–785.
29. Stinson JN, McGrath PJ, Yamada JT. Clinical trials in the Journal of Pediatric Psychology: Applying the CONSORT statement. *Journal of Pediatric Psychology*. 2003;28:159-167.
30. Boutron I, John P, Torgerson DJ. Reporting methodological items in randomized experiments in political science. *The ANNALS of the American Academy of Political and Social Science*. 2010;628(112):112-131.
31. McGrath PJ, Stinson J, Davidson K. Commentary: The Journal of Pediatric Psychology should adopt the CONSORT statement as a way of improving the evidence base in pediatric psychology. *Journal of Pediatric Psychology*. 2003;28:169-171.
32. Gill CE. Missing links: How descriptive validity impacts the policy relevance of randomized controlled trials in criminology. *Journal of Experimental Criminology*. 2011;7(3):201-224.
33. Grant S, Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo-Wilson E. Developing a reporting guideline for social and psychological intervention trials. *Research on Social Work Practice*. 2013;23(6):595-602.
34. Montgomery P, Grant S, Hopewell S, et al. Protocol for CONSORT-SPI: An Extension for Social and Psychological Interventions. *Implementation Science*. 2013;8:99.
35. Oakley A, Strange V, Bonell C, Allen E, Stephenson J. Process evaluation in randomised controlled trials of complex interventions. *BMJ*. 2006;332:413-416.

36. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.

## Publications and Presentations Related to CONSORT-SPI to Date

### Publications

**Grant S.** Book review: Reporting Research in Psychology: How to Meet Journal Article Reporting Standards. *Research on Social Work Practice*. 2014;24(2):253-255.

**Grant SP,** Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.

**Grant S,** Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo-Wilson E. Developing a reporting guideline for social and psychological intervention trials. *Research on Social Work Practice*. 2013;23(6):595-602.

\*This manuscript is co-published in:

- *Journal of Experimental Criminology*. 2013;9:355-367.
- *British Journal of Psychiatry*. 2013;203: 250-254.
- *American Journal of Public Health*. 2013;103(10):1741-1746.
- *Trials*. 2013;14:242.
- *British Journal of Social Work*. 2013;43(5):1024-1038.

**Grant S,** Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo- Wilson E. Letter to the Editor: New guidelines are needed to improve the reporting of trials in addiction sciences. *Addiction*. 2013;108:1687-1688.

Gardner F, Mayo-Wilson E, Montgomery P, ... **Grant S.** Editorial Perspective: The need for new guidelines to improve the reporting of trials in child and adolescent mental health. *Journal of Child Psychology and Psychiatry*. 2013;54(7):810–812.

Montgomery P, **Grant S,** Hopewell S, et al. Protocol for CONSORT-SPI: An Extension for Social and Psychological Interventions. *Implementation Science*. 2013;8:99.

**Grant S,** Montgomery P, Mayo-Wilson E. Development of a CONSORT extension for interventions in public health and related disciplines. *Lancet*. 2012;380:S14.

Spreckelsen T, **Grant SP,** Montgomery P. Letters: Additional requirements for complex interventions. *BMJ*. 2012;345:e8003.

## **Presentations**

**Grant S.** Developing a CONSORT extension for social and psychological interventions. Paper presented at: Green Templeton College, University of Oxford. 2013, May: Oxford, UK.

Montgomery P, **Grant S**, Mayo-Wilson E. Reporting quality of complex social intervention trials: A CONSORT extension for complex social interventions. Paper presented at: 17<sup>th</sup> annual Society for Social Work and Research (SSWR) Conference. 2013, January: San Diego, USA.

**Grant S**, Montgomery P, Mayo-Wilson E. Development of a CONSORT extension for interventions in public health and related disciplines. Paper presented at: Royal Society of Medicine's Public Health Science: A National Conference Dedicated to New Research in Public Health. 2012, November: London, UK.

**Grant SP**, Montgomery P. Development of a CONSORT Extension for complex social interventions. Meeting at: 20<sup>th</sup> annual Cochrane Colloquium. 2012, October: Auckland, New Zealand.

Montgomery P, **Grant SP**. Development of a CONSORT extension for randomised controlled trials evaluating complex behavioural change and psychological interventions. Meeting at: 19<sup>th</sup> annual Cochrane Colloquium. 2011, October; Madrid, Spain.

Montgomery P, **Grant SP**. Will you CONSORT with me? Paper presented at: Centre for Evidence-Based Crime Policy and Campbell Collaboration Joint Symposium on Evidence-Based Policy. 2011, August: Washington DC, USA.

Montgomery P, Mayo-Wilson E, **Grant SP**. Will you CONSORT with me? Improving reporting quality to improve outcomes for people and organizations. Poster presented at: 1<sup>st</sup> annual Global Implementation Conference. 2011, August: Washington DC, USA.

## **Chapter 1: Randomised Controlled Trials of Social and Psychological Interventions**

*NB: Sections of this chapter have been published in an alternative form in  
Research on Social Work Practice<sup>1</sup>*

### **1. Introduction**

This chapter will discuss the use of randomised controlled trials (RCTs) to evaluate social and psychological interventions. The first section of this chapter consists of a brief introduction to scientific research methods for investigating causal relationships; the increasing use of RCTs to evaluate medical interventions with the emergence of evidence-based medicine (EBM); and the subsequent prominence of RCTs in areas outside of medicine under the more general evidence-based practice (EBP) paradigm. Next, it will detail modifications to conventional design features of RCTs when used to evaluate social and psychological interventions. These design features aim to account for the primacy of intervention theory, processes of intervention implementation, and trial context when examining how social and psychological interventions work, for whom, and under what conditions.<sup>2,3</sup> Lastly, this chapter will discuss the importance of reporting standards for health research, and will argue that a new reporting guideline may be needed for social and psychological intervention RCTs.

### **2. Scientific Approaches to Establishing Causal Relationships**

A major goal of intervention research is the detection of dependable causal relationships between an intervention and outcomes of interest. A variety of scientific approaches for investigating such causal relationships exist. Currently, experimental approaches are often promoted—when feasible and appropriate—as the most reliable method for establishing causality.<sup>4</sup> This section will briefly trace the progression of

scientific experiments from the conventional RCT in medicine to the current “best-practice” of mixed-method RCTs for evaluating social and psychological interventions.<sup>5</sup>

### ***2.1 Scientific Experiments***

Experiments, in the general sense of examining the world for the sake of discovery, are part and parcel with human nature. *Scientific* experimentation in particular involves investigative procedures that are orderly and systematic, using empirical methods to test models or hypotheses that aim to explain particular phenomena of interest.<sup>6</sup> To make the results of experiments more trustworthy, scientists employ a variety of design features that address factors extraneous to the model or hypothesis but could still impact the results.

A noteworthy design feature of some experiments that seek to establish causal relationships between an independent and dependent variable is the use of a comparison group.<sup>7</sup> The trustworthiness of results from experiments using this “counterfactual” approach largely depends on the ability of the researcher to do three crucial factors.

The first factor is for the researcher to be able to construct two samples that are identical in every way except for the presence of the independent variable, in order to rule out alternative explanations for observed effects.<sup>4</sup> Should the “counterfactual” sample or comparison group be identical to the experimental group except for the independent variable, the experiment would then seek to establish whether the presence of the independent variable leads to differences in outcomes between the two samples.

The second factor is to manipulate the uniform presence of an independent variable. The ability to assign the independent variable to the experimental sample at a known point in time establishes the temporal precedence of the independent variable before any changes in the dependent variable or outcome of interest between the otherwise identical groups. This temporal precedence aims to further support inferences that the independent variable

caused an observed effect.

The third factor is to control the context of the experiment so that all other potentially confounding variables are held constant—and, ideally, are of known values. The researcher could also systematically alter contextual variables of interest across experimental tests to see if these variables change any effects of the independent variable on the outcomes of interest.

These design features increase the experiment's internal validity: i.e., the extent to which the conclusion of a causal relationship between the independent and dependent variables is warranted.<sup>7</sup> Many scientific experiments have derived their strength from any or all of these methodological procedures. However, as these procedures gained prominence in medical experiments, more robust techniques were needed to create counterfactual conditions and to rule out confounding factors when investigating the effectiveness of interventions in or on human biological systems.

## ***2.2 Randomised Clinical Trials in Medicine***

The advancement of scientific experimental methods in clinical medicine has its roots in R.A. Fisher's and Austin Bradford Hill's seminal contributions to the design of randomised field trials in the early 20<sup>th</sup> century. During the 1930s, Fisher showed how randomisation provides a theoretical underpinning for tests of statistical significance in experiments via unbiased random selection of experimental and comparison samples. On the basis of Fisher's work, randomisation was later promoted for use in clinical trials of medical treatments by Austin Bradford Hill in the 1940s.<sup>8</sup> Influenced by the work of Fisher and other scientific predecessors, Hill provided nine considerations for medical researchers to assess the confidence in claims that an observed association in fact was likely to involve a causal relationship. These considerations included the strength of association between the

purported cause and effect, the temporal precedence of the cause, a dose-response relationship (i.e., changes in the effect that systematically correspond to differing levels of exposure to the cause), and evidence from formal scientific experimentation.<sup>9</sup> These considerations, often known as the Bradford Hill criteria or Hill's criteria for causation, have had an enormous influence on experimental approaches to separate causal from non-causal explanations of observed associations in epidemiological and clinical research.<sup>10</sup>

As the Bradford Hill criteria increased in prominence in medical research, RCTs generally came to be considered the most valid and reliable research method for estimating the effectiveness of medical interventions.<sup>11</sup> RCTs achieved this "gold standard" status because successful random allocation of participants to different interventions has been argued to be the most secure method of creating comparison groups that on average are similar to the experimental group except for receipt of the medical intervention. Put more methodologically, randomly assigning participants to either receive or not receive the intervention prevents the selection bias in which systematic, baseline differences in confounding variables exist between the experimental and comparison groups.<sup>12</sup> If such a selection bias occurs, any post-treatment differences in outcomes between the two trial groups could be attributed to systematic differences in prognostic baseline variables rather than the intervention of interest. However, if researchers are successful in randomly allocating (a sufficient number of) participants to each of the trial groups, then any baseline differences between the two groups in traits, exogenous experiences, and unknown or unmeasured confounding variables are not purported to be systematic but rather due to chance alone.<sup>13</sup> The only systematic difference then between trial groups is the allocated intervention. Thus, any post-treatment group differences in trial outcomes can be confidently attributed to the different interventions that each group received.

Other methodologies besides RCTs do exist for evaluating interventions. However,

those who favour RCTs do so because of their high assurance of internal validity by protecting against systematic baseline differences in prognostic variables between an experimental and comparison group.<sup>12,14</sup> Other research designs that employ a counterfactual must use a greater number of statistical assumptions in order to reduce this form of selection bias, though none of these designs theoretically can account for unmeasured confounding variables as adequately as random assignment.<sup>15</sup>

### ***2.3 Evidence-Based Medicine***

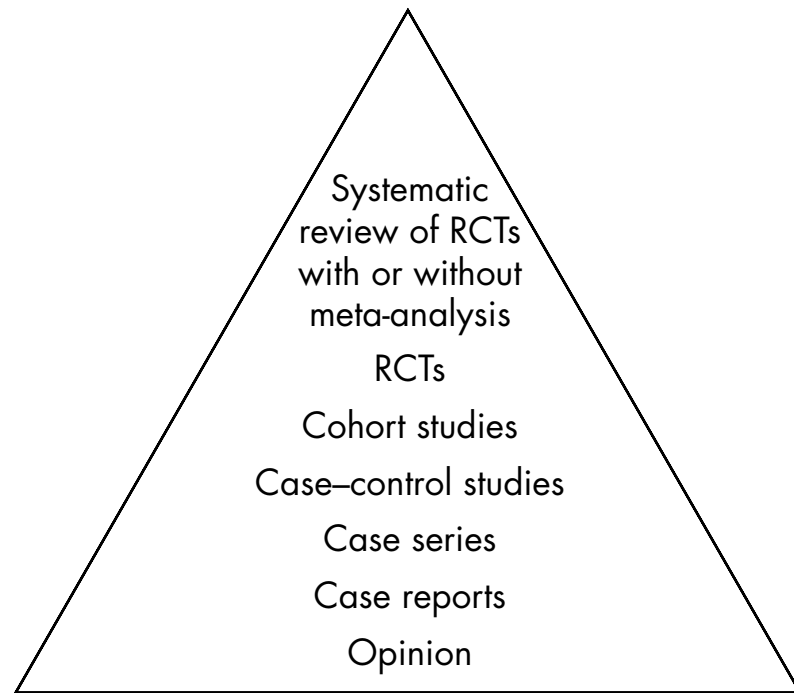
The work of Fisher and Bradford Hill laid foundations for the evidence-based medicine (EBM) paradigm, the origins of which trace back to a series of lectures (published in 1972) by the epidemiologist Archie Cochrane titled *Effectiveness and efficiency: Random reflections on health services*.<sup>16</sup> Cochrane argued that too many medical interventions were of dubious or unknown safety and effectiveness, with the potential to cause massive harm at both individual and population levels, through iatrogenic injury, waste of resources, and failure to take up more effective treatments.<sup>17</sup> Cochrane then proposed that interventions should be evaluated rigorously and systematically, using unbiased methods of evaluation like the RCT. Cochrane also proposed that individuals in the medical professions should continuously update their knowledge of intervention effectiveness. Cochrane's approach demonstrated a strong ethical imperative to do no harm and to do the most good possible for one's patients by using research evidence to inform clinical decisions.<sup>11</sup> In 1991, the term "evidence-based medicine" emerged, later defined as the integration of the best available research evidence with clinical expertise and patient values in clinical decision-making.<sup>18,19</sup>

## ***2.4 The Evidence Hierarchy of EBM***

Prior to EBM, three methods for determining intervention effectiveness seemed to compete for authority in clinical decision-making: (1) comparing the outcomes of patients receiving different treatments for the same condition, (2) specifying the underlying mechanisms of a treatment, and (3) authoritative statements or received wisdom from experts with clinical experience in the area.<sup>20</sup> Largely influenced by Fisher, Bradford Hill, and Cochrane, proponents of EBM strongly supported the “direct comparison of treatments” approach, though they incorporated all three methods into the EBM “hierarchy of evidence” (see Figure 1).<sup>20</sup> This hierarchy purports to delineate which methods are least likely to be at risk of bias when aiming to reliably and validly estimate the effects of an intervention, or whether an intervention caused outcomes of interest.<sup>21</sup>

Of all comparative treatment methods, proponents established RCTs—or systematic reviews of RCTs where possible—at the top of this hierarchy. The hierarchy is partly based on a strand of empirical research demonstrating that several widely used treatments with supportive evidence from the “lower” methods on the hierarchy were actually shown to be ineffective or harmful when tested more rigorously via “higher” methods on the hierarchy, such as RCTs. Examples of traditional medical interventions subsequently found to be harmful include: providing oxygen to premature babies, leading to more blindness in these babies; giving anti-arrhythmic drugs to post-heart attack patients, leading to more deaths; and administering synthetic hormones to prevent miscarriage, leading to female daughters developing rare vaginal cancers and other diseases.<sup>11</sup> Nonetheless, it is interesting to note that, even in EBM circles, there has been some controversy about proposed methods to assess intervention effectiveness, namely the privileging of RCTs over other methods on first principles.<sup>17</sup> These debates foreshadow disputes about applying the EBM hierarchy to social and psychological interventions.

**Figure 1. EBM Hierarchy of Evidence for Determining Intervention Effectiveness<sup>22</sup>**



### **3. The Evidence-Based Practice Movement**

#### ***3.1 Experimentation in the Social and Behavioural Sciences***

Precedents of experimentation within the social and behavioural sciences have laid the groundwork for the use of RCTs to evaluate social and psychological interventions.<sup>23</sup> Though some form of randomised experiments appeared in psychology and education prior to clinical medicine, many disciplines in the social and behavioural sciences began using randomised trials to examine intervention effectiveness after Fisher's seminal work. Early examples of such RCTs range from large experiments in education conducted by Walters in 1931 and 1932<sup>24</sup> to direct low-income housing cash assistance in the US in the 1970s.<sup>25</sup> Within recent decades, numerous examples demonstrate the feasibility of conducting social and psychological intervention RCTs, as evidenced, for example, by reviews in the Cochrane and Campbell Collaborations.<sup>26</sup> RCTs have gained particular prominence as of late due to the emergence of the Evidence-Based Practice movement.

### ***3.2 Emergence of Evidence-Based Practice***

As EBM gained traction in clinical medicine, other researchers gradually began to notice the pertinence of the paradigm's principles to decision-making in their own areas. As a result, the primary tenets of EBM were translated into a more general paradigm known as "evidence-based practice" (EBP) to allow its applicability to areas outside of medicine.<sup>23</sup> Expanding from medicine, EBP was defined as a model of decision-making in which the best available scientific evidence from well-designed and conducted studies should be integrated with practitioners' or policy-makers' expertise and clients' preferences to inform practice and policy for a given discipline.<sup>18</sup>

As in EBM, EBP specifically highlights the use of *reliable* research data to support decisions. Thus, the paradigm still hinges on the ability to stay up-to-date with the most current research, to critically appraise studies, and to integrate new findings with previous knowledge.<sup>27</sup> Granted, EBP posits that the integration of high quality published evidence *in addition to* provider expertise and the views of intervention recipients will lead to the best policy and practice decisions. Nonetheless, for this thesis, the DPhil candidate wishes to highlight that the paradigm gained momentum under the auspices of including rigorous, systematic insights from *research evidence* into practice and policy decision-making. Moreover, in adopting the EBM hierarchy of evidence, the more general EBP paradigm has also tended to axiomatically privilege RCTs (and systematic reviews of them) for identifying and promoting the interventions that work and eliminating the interventions that are ineffective or harmful.

### ***3.3 EBP for Social and Psychological Interventions***

In the late 1990s, a number of commentaries pointed to the lack of robust evidence to support social and mental health policies in the UK and North America.<sup>21</sup> At the time, and

likely still today, there were common misconceptions by some that social and psychological interventions do not have the capacity to do harm, so robust evidence investigating their effects was not needed. However, as noted earlier, even interventions with the best of intentions may not work, and sometimes may actually be harmful in practice.<sup>11,28</sup> There are numerous examples of social policies and interventions that have no benefit or have actually done harm, such as Scared Straight and Boot Camp programmes for juvenile delinquents that actually increase recidivism rates.<sup>21,29</sup> As awareness of the potential ineffectiveness or harm of social and psychological interventions rose, the use of reliable evaluation methodologies became of practical, ethical, and scientific importance.<sup>30</sup> Examples include the standards of the American Psychological Association Task Force on Empirically Supported Treatments,<sup>31</sup> the Substance Abuse and Mental Health Services Administration's National Registry of Evidence-Based Programs and Practices,<sup>32</sup> and the Institute of Education Sciences What Works Clearinghouse,<sup>33</sup> amongst other groups.

As EBP spread to social and psychological intervention research, a simultaneously attractive and contentious component of this new paradigm was its evidence hierarchy. The attractiveness of the hierarchy was that, for conflicting professional opinions, EBP offers an adjudication process via standards for the evidentiary status of research studies. Many expert policy makers and practitioners say they are "certain" about the effectiveness of differing practices, yet provide discordant opinions amongst themselves.<sup>11</sup> RCTs and meta-analyses of interventions can help arbitrate these conflicting expert opinions, for the methodology of these studies and the statistics applied to examine them help to rule out the possibility of attributing effects of bias, confounders, and chance to the intervention.<sup>34</sup> Rigorous critical appraisal of research using explicitly agreed upon standards and skills for appraising a study's validity, impact, and applicability<sup>35,36</sup> can therefore inhibit inflated or inaccurate claims of effectiveness by policy-makers, providers, and service users.

Considering the exponential increase in accessible, practice-relevant research, a greater incorporation of scientific evidence in decision-making has the potential to increase effectiveness of programmes and enhance a profession's credibility.

#### **4. Evaluating Social and Psychological Interventions with RCTs**

There is still considerable controversy as to whether the RCT is indeed a suitable design—let alone the best design—for evaluating social and psychological interventions,<sup>30,37</sup> or at least whether it should always be privileged over other methods based on foundational assumptions.<sup>38</sup> Scepticism is in part due to misconceptions that RCTs have a solely biomedical origin entirely inapplicable to this area.<sup>39</sup> Such conceptions often lead to doubts about the ability of RCTs to make causal inferences in the more complex world of mental processes, the social life, and human behaviour.<sup>40</sup> Common critiques include the inappropriateness of RCTs to certain social and psychological phenomena, the infeasibility of maintaining strict adherence to trial protocols, and ethical concerns about randomising participants.<sup>37</sup> For example, some argue that the environment of “human and social services” rarely remains stable long enough to conduct an experiment from which confident conclusions can be made and applied to later situations.<sup>41</sup>

Despite such controversy, the RCT is seen by many as a best-practice for establishing the effectiveness of social and psychological interventions.<sup>42</sup> However, there are many possible aspects of social and psychological interventions that make evaluating them with RCTs particularly susceptible to improper design and execution. These challenges leave some researchers, who nevertheless agree in principle with using RCTs, to consider using rigorous non-randomised study designs instead.<sup>43</sup>

To address concerns about the suitability of RCTs in this area, social and behavioural scientists have started to adapt and integrate diverse yet long-standing scientific

approaches for investigating causality into RCT designs. These researchers have sought to modify RCTs with design features that better meet their methodological needs, while respecting the distinct branches of philosophical thought underlying these newly adopted approaches.<sup>6</sup> The DPhil candidate shares the view that RCTs should be part of the “evaluation toolkit” for social and psychological interventions, though the characteristics of these interventions provide challenges that need to be addressed in trial design, execution, analysis, and reporting.

#### ***4.1 Characteristics of Social and Psychological Interventions***

To identify useful design features, scientists have highlighted key characteristics of social and psychological interventions (see Box 1). Namely, the mental processes and social phenomena targeted by these interventions are inextricable with human agency: i.e., the capacity of those involved in providing and receiving the intervention to make choices and enact them within the bounds of surrounding physical and societal structures.<sup>44</sup> This dependence on working through human agency can lead to tremendous differential effects in the outcomes being pursued by an intervention. Consequently, social and psychological interventions are generally not single, static treatments uniformly implemented, but are often conceptualised as a “programme offer”.<sup>45</sup>

Under this conceptualisation, implementation of the intervention protocol involves a continuous, interactive *process* of providers and participants in a given context who respond (potentially differentially) to the techniques, activities, services, and so on that constitute the intervention on offer.<sup>44</sup> Ideally, then, social and psychological intervention researchers must take account of the human agency and intentionality of providers and participants in their theory of change (i.e., the hypothesis of how the intervention is supposed to work), the choice of techniques or activities employed in these interventions,

### Box 1: Features of Social and Psychological Interventions<sup>46,47</sup>

Accounting for human agency and intentionality in theory of change leads to:

- Social and psychological processes as malleable mechanisms targeted by the intervention, which works through changes in an individual's or group's cognitions, emotions, attitudes, behaviours, relationships, norms, culture, salient environments, and so on.
- use of behavioural change, psychological, and structural intervention techniques.
- difficult to measure "latent" constructs.
- delivery in "hard-to-control" environments.

Important aspects of interventions to consider during development, evaluation, and implementation:

- multiple, interacting intervention components.
- the number and difficulty of behaviours required by those delivering and receiving the intervention.
- the number of groups or organisational levels targeted by the intervention.
- various outcomes on multiple levels.
- the degree of flexibility or tailoring of the intervention permitted.

and the role of context when designing evaluations that explain intervention effects and any sources of variation in these effects.<sup>37,48</sup>

#### 4.1.1 *Intervention Theory and Mechanisms of Action*

One key issue of social and psychological intervention research is gaining an understanding of the nature of targeted mechanisms and how intervention techniques influence them within a given context. It may prove difficult to precisely assess changes in mechanisms of action and how these changes affect outcomes of interest, particularly when mechanisms are multiple or when they have not been clearly specified prior to the delivery of an intervention. Theories provide sound explanations for the various ways that participants will respond to intervention techniques in a given context to produce certain outcomes.<sup>49</sup> A clear theory of change is therefore crucial for elucidating the consequences of social and psychological interventions to stakeholders.

#### *4.1.2 Intervention Implementation*

Social and psychological interventions techniques are often not completely uniform in their delivery and receipt, if not quite heterogeneous in their actual implementation. The “active ingredients” of social and psychological interventions involve various techniques that target intrapersonal processes (e.g., cognitions, affections), interpersonal processes (e.g., social relations), and structural processes (e.g., physical layout of a community). Several complexities consequently abound in implementing these interventions, due to the number of intervention components, the potential of components to have a synergistic effect, and the dependency of active ingredients on human agency.<sup>49</sup> Key to understanding what the human agents do during implementation of an intervention are the functions of intervention techniques, the new possibilities or capabilities that are now obtainable because of the intervention, and the potential for agents to mobilise social and structural resources to realise these new possibilities.<sup>44</sup>

#### *4.1.3 Contextual Dependence of Intervention Effects*

Social and psychological interventions are often delivered in hard-to-control settings with many potentially confounding influences. Examples include health care settings (e.g., primary, secondary, or tertiary care); community settings (e.g., schools and social service agencies); and correctional facilities. The way in which social and psychological interventions influence targeted mechanisms to obtain desired effects may be dependent on and interact with aspects of a given context.<sup>7</sup> A common, useful conceptualisation of causality for elucidating this principle of contextual dependence is the “INUS condition”: i.e., social and psychological interventions are not themselves the sole cause of an outcome, but rather are insufficient (I) but non-redundant (N) parts of a condition that itself is an unnecessary (U) yet sufficient (S) condition for the outcome to occur.<sup>7</sup> Different

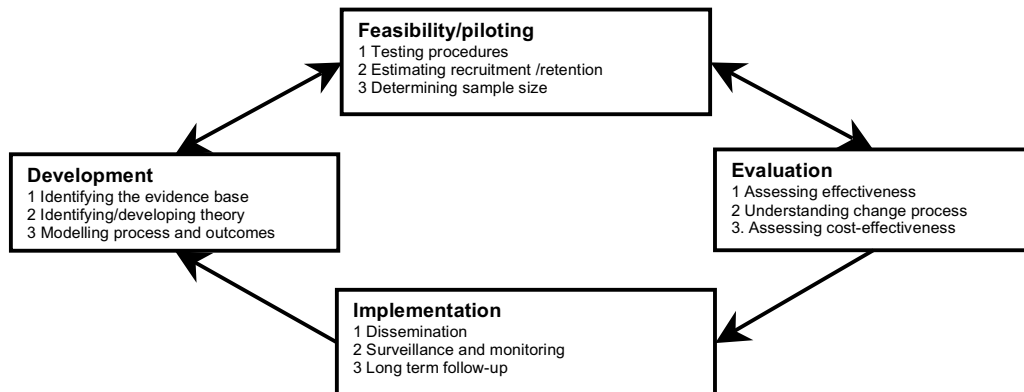
underlying social, institutional, psychological, and physical structures may yield different probabilistic causal-relations between an intervention and observed outcomes. This potential for contextual dependence leads to problems with judging the applicability of previous research to the trial setting and population, as well as the applicability of trial results to other contexts.

#### ***4.2 Design Features for RCTs of Social and Psychological Interventions***

The above characteristics of social and psychological interventions pose significant challenges to designing RCTs that have the ability to estimate and explain intervention effects. A useful structure for conceptualising features of social and psychological interventions for the purposes of evaluation is the *MRC Framework for the development and evaluation of complex interventions*<sup>50</sup> (see Figure 2). Features of complex interventions as defined by this framework include (1) the solitary as well as the possible synergistic effects of multiple, interacting intervention components; (2) the number and difficulty of behaviours required by those delivering and receiving the intervention; (3) the number of groups or organisational levels targeted by the intervention; (4) various outcomes of these interventions that exist on multiple levels (e.g., individual, group, community); and (5) the permitted degree of flexibility or tailoring of the intervention.<sup>50</sup>

Rather than seeing these as dichotomous features that are either present or not, it is more useful to see these features on a continuum, where each is more or less present depending on the intervention being evaluated. The more each feature is present, the more complex an intervention may be, and the more one might expect differential responses to the programme offer that need to be investigated. Such complexities—which are often found in social and psychological interventions—complicate the ability of RCTs to explain how and why an intervention works, for whom, and under what conditions.

**Figure 2. Key Elements of the Updated MRC Framework for the Development and Evaluation of Complex Interventions<sup>50</sup>**



To address these challenges, social and psychological intervention researchers have developed trial design features that allow more in-depth investigation into what is happening in each experimental condition during a trial. Namely, delineating intervention theory and embedding process evaluations within an RCT help to clarify how mechanisms are operating as well as the influence of context on the propensity for an intervention to produce desired effects.<sup>4,46</sup> Combined with the traditional aspects of an RCT (for instance, creating a comparison group via random assignment), researchers can aim to not only identify what interventions “work,” but also for whom, how, and under what conditions. Such mixed-methods research is indeed advocated in the MRC Framework<sup>47,51</sup> as well as draft guidance on process evaluations in trials of complex interventions,<sup>52</sup> in order to increase confidence in causal claims and interpretations. Many design features applied within social and psychological intervention RCTs are examples of such solutions for strengthening the internal and external validity of trials evaluating complex interventions.<sup>4,7,24</sup>

#### *4.2.1 RCT Design Features Related to Internal Validity*

A major challenge for the internal validity of these RCTs is that social and psychological interventions consist of several interacting components that have the potential to simultaneously target individual, community, and environmental constructs and often target outcomes that cannot be directly observed (e.g., beliefs, knowledge, and emotions).<sup>47</sup> To address this challenge, the use of theory and conceptual frameworks for intervention design and implementation is highlighted considerably in scientific literature. Researchers are called to develop a clear theory of change to understand the mediators, moderators, and contextual dependencies of these interventions, so that weak links in the causal chain can be identified, tested, and modified.<sup>53</sup> Such intervention trials tend to incorporate theoretical perspectives in the design and implementation of both the intervention and the study itself. This use of theory is important for building on existing scientific knowledge from a variety of disciplines, facilitating coherent syntheses of evidence and theory after a trial, and implementing an intervention in a context different than its evaluation.<sup>54</sup>

Researchers must also consider the validity and reliability of measures of latent constructs in the experimental environment.<sup>55,56</sup> Psychosocial researchers often investigate constructs—such as attitudes and mental states—that cannot be directly observed due to their intangible nature. Instead, these constructs are often indirectly measured via self-report questionnaires or researcher observation of participant behaviours related to these constructs. With regards to measurement, “validity” is an assessment of whether an instrument measures what it aims to measure, whereas reliability represents the reproducibility and consistency of an instrument over time.<sup>6</sup> Both are important to trustworthy research findings. There are several potential threats to the reliability and validity of such indirect measures in the experimental environment, depending on how

they were designed and implemented, that can lead to systematic deviations from the true value of the construct. For example, a review of 300 RCTs evaluating interventions for schizophrenia found that outcomes measured via unpublished scales, which typically had no supporting data on the validity or reliability of the scale, were more likely to report that the intervention was superior to control; the review authors concluded that one-third of claims of treatment superiority would not have been made if published scales had been used instead.<sup>57</sup>

Important aspects of self-report measures include their psychometric properties and measurement format. Psychometric properties of reliability and validity for an instrument are assessed through a series of defined tests on the population group for whom the instrument is intended.<sup>6</sup> Researchers should take into account the differences between the sample(s) used to develop the measure and the sample of their intervention study in order to appraise how reliable and valid the instrument is likely to be for their study. Regarding measurement format, psychosocial researchers can attempt to overcome reliability issues by utilising several sources (e.g., participants' spouses, family members) and measures (e.g., questionnaires, observation) to measure a construct.

Researchers must also reflect on which statistical analyses are appropriate for these data, as multi-level data of latent constructs can often been analysed in several ways.<sup>55,56</sup> In addition, researchers should deliberate on how to detect adverse events, as harms of social and psychological interventions are not always as direct or obvious as mortality or morbidities in health care research.<sup>24</sup> Depending on the discipline, adverse events could be construed as criminal acts, truancy, or damages to relationships, amongst other psychosocial outcomes.

Another issue relates to the blinding of participants and providers: i.e., masking the allocated intervention during a trial. Such blinding is generally not possible in social and

psychological intervention research because both providers and participants are aware of what interventions participants receive. Awareness of intervention assignment is often in fact necessary for proper implementation of the intervention. Consequently, researchers must consider how this awareness can lead to performance biases and demand effects by providers and participants that influence measured outcomes. Moreover, while it may be impossible to blind providers and participants, blinding researchers during data collection and analysis is often possible and can address concerns related to internal validity and bias. Otherwise, unblinded outcome assessment is another potential source of bias affecting trial results, especially considering that many psychosocial outcomes are subjectively assessed.<sup>58</sup>

#### *4.2.2 RCT Design Features Related to External Validity*

For those seeking to apply social and psychological intervention research, threats to external validity may be as important as threats to internal validity. Contextual details are significant for these stakeholders, especially considering that trials often take place in settings and with populations that differ from real-world situations due to population selection biases.<sup>59</sup> Lack of information relevant to external validity may prevent practitioners or policy makers from using evidence appropriately to inform decision making; yet, existing guidelines do not adequately explain how authors should describe (a) how interventions work, (b) for whom, and (c) under what conditions.<sup>60</sup> Greater attention to a trial's contextual features could address criticisms that RCTs have little generalisability to real-world settings, as contextual information is necessary to compare the effectiveness of an intervention across time and place.<sup>61</sup>

One area of external validity that receives great attention in social and psychological intervention research is implementation fidelity and replication.

Implementation fidelity refers to the degree that interventions were delivered by practitioners and taken up by participants as designed.<sup>62</sup> It is useful for authors to examine the key components of interventions, how those components could be delivered, and how they relate to the outcomes selected. Moreover, interventions are rarely implemented exactly as designed, and social and psychological interventions may be designed to be implemented with some flexibility, in order to accommodate differences across participants,<sup>63</sup> so it is important to determine how interventions were actually delivered by providers and actually received by participants.<sup>64</sup> Particularly for social and psychological interventions, the integrity of implementing the intended functions and processes of the intervention are essential to understand.<sup>63</sup> As RCTs of a particular intervention can yield different relative effects depending on the nature of comparison groups, information about delivery and uptake should be assessed for all trial arms.<sup>65</sup>

Psychosocial researchers have expressed concern over the dearth of RCTs that adequately consider implementation and fidelity.<sup>66</sup> Many reports neither contain sufficient information about the interventions tested nor reference treatment manuals.<sup>67</sup> Developing logic models—as described in the MRC Framework for Complex Interventions<sup>50</sup>—or theories of change can help elucidate links in causal chains that can be tested, identify important mediators and moderators, and facilitate syntheses in reviews.<sup>68</sup> Significant detail should be gathered about provider fidelity to protocol, participant compliance with core components of an intervention, different formats of delivering the intervention, and the training and supervision of the providers for *all* trial arms in social and psychological intervention RCTs.<sup>65</sup> Measures of implementation fidelity are important for understanding what exactly occurred in a trial and how to successfully transport an intervention from research to real-world settings.

Other aspects of external validity relate to the comparability of study samples and

settings to real-world populations and environments. An intervention that works for one group of people may not work for people living in different cultures or physical spaces, or it may not work for people with slightly different problems and comorbidities. Participants in RCTs of social and psychological intervention are often recruited outside of routine practice settings via processes that differ from routine services.<sup>59</sup> Random sampling is also often seldom used. Enrolling in an RCT can be a complex process that affects the measured and unmeasured characteristics of participants, and recruitment may differ from how users normally access interventions. Well-designed RCTs (with adequate resources) look to assess characteristics of all participants (volunteers, those who enrolled, and those who completed) in sufficient detail to understand the comparability of the study sample to targeted populations in every-day services.<sup>59,69,70</sup> These details about recruitment processes and participant selection are important for understanding the transportability of experiments to target populations.

Finally, given that these interventions often occur in hard-to-control environments, trialists should seek to examine factors of RCT contexts that are believed to support, attenuate, or frustrate observed effects.<sup>71</sup> Adaptations of social and psychological interventions to contextual features are inevitable, so researchers need to identify the mechanisms targeted by interventions, necessary intervention components as well as required auxiliaries components, and barriers to successful intervention delivery and uptake. Researchers need to take the characteristics of both trial sites and target communities into consideration so that research can help to identify whether interventions have a stable capacity to be effective in real-world settings.<sup>61</sup> Several aspects of setting and implementation may be important to consider for the applicability of trials to other contexts, such as administrative support, staff training and supervision, organisational resources, the wider service system, and concurrent political or social events.<sup>49,72-74</sup>

Flexibility in the implementation of social and psychological interventions may require detailed information about participant socioeconomic characteristics and programme preferences.<sup>47,53,75,76</sup> As such, interventions may also differ across groups of different social or socioeconomic positions, so any equity considerations should be addressed explicitly.<sup>77,78</sup>

## **5. Reporting Standards and Guidelines for Research**

### ***5.1 The Need for Transparent Reporting***

Growing attentiveness to the quality of *conducting* RCTs has led to a rising interest in the quality of *reporting* RCTs in journal articles and publications.<sup>79</sup> Reporting the findings of an RCT is a stage of intervention research inseparably linked to trial design and conduct. Because RCTs can be inadequately designed or executed, the ability of providers to ethically and appropriately apply research findings depends on the information presented in trial reports. The internal and external validity of a trial can only be illuminated through transparent reporting: i.e., the explicit description of a study's logic and key features from study design to data collection, analysis, and interpretation.<sup>59</sup> RCT reports need to be sufficiently transparent in order to facilitate critical appraisal, which is vital to interpret the legitimacy and relevance of RCT findings to particular "real-world" contexts.<sup>80</sup> Incomplete and unclear reporting hinders the appraisal of RCTs, leaving readers susceptible to implementing biased or irrelevant research findings in their professional settings. When detailed information about study conduct is not reported, or when it is not reported well, the link between research and practice is broken. In this way, scarce resources are routinely used to sub-optimal effect.<sup>81</sup> Thus, to have its intended impact, the methods for reporting research must be as rigorous as those for conducting it.

Proper reporting is especially essential for synthesising RCTs into high quality

systematic reviews, which are increasingly used to inform practice guidelines and public policy.<sup>49</sup> When reported clearly and completely, good research can be included in systematic reviews and practice guidelines, leading to better service outcomes. Systematic reviewers require a substantial amount of information from trial reports to (a) determine which RCTs to include in a review, (b) examine biases within included studies, (c) pool study data for meta-analysis, and (d) clarify the generalisability of results.<sup>82</sup> Important details include information on the randomisation procedure, descriptions of interventions as they were actually delivered in individual trials, and data on all outcomes in a format that is amenable to meta-analysis.<sup>83</sup> As insufficient information hinders the robustness of a review, poor reporting in RCT publications can have serious consequences for practice, research, policy-making, and ultimately intervention recipients.

For example, Multisystemic Therapy (MST) is an intensive, complex psychosocial intervention for chronic juvenile offenders. This intervention has been given over £55 million in research grants by U.K, U.S., and Canadian funding agencies.<sup>84</sup> It is being implemented in 13 countries to over 23,000 families per year, and public finances are progressively covering its £3,400 annual cost per family.<sup>84</sup> Despite numerous RCTs on the intervention, however, Littell and colleagues' systematic review on MST concludes that evidence of its effectiveness is uncertain.<sup>85</sup> Though many factors contribute to these inconclusive findings, the authors explicitly state that partial responsibility lies with inadequate reports of (a) randomisation procedures, (b) the number of participants assigned to each trial group, and (c) other "errors of omission."<sup>85</sup> Incomplete and unclear information in RCT publications on MST left the review authors to make several assumptions about important trial details, such as the number of participants in each analysis and the methodological quality of included studies.<sup>85</sup> These study details bear direct influence on meta-analytic effect estimates.<sup>86</sup> If the actual details significantly

differed from the reviewers' assumptions, the meta-analysis may have provided a different and possibly more conclusive effect estimate.<sup>85</sup> Instead, millions in research funds have provided potentially speculative conclusions. This systematic review is but one example demonstrating that the quality of *reporting* research is as important for evidence-based decision-making as the quality of *conducting* research.

### ***5.2 Evidence of Poor Reporting***

Academic journals are the most common outlet for researchers to report results of RCTs, and scientific publications in academic journals are thus the primary means of distributing RCT findings to the wider research community. As these publications are often the only concrete, public records of RCTs, transparent, complete, and accurate journal articles are essential for researchers, practitioners, and policy-makers to meaningfully utilise trial findings.<sup>83</sup> However, partly as a result of the complexities of social and psychological intervention RCTs, there is significant evidence suggesting poor reporting of these trials: i.e., publications often fail to include essential information for critical appraisal. Poor reporting of social and psychological interventions has been identified in criminology,<sup>82</sup> education,<sup>87</sup> psychology,<sup>75</sup> public health,<sup>88</sup> and social work.<sup>66</sup>

Many important aspects of RCTs related to internal validity remain poorly reported in social and psychological intervention trial publications. For example, a systematic review of 62 RCTs in criminology revealed consistent poor reporting of outcome measures, data, and statistical analyses.<sup>82</sup> Without complete information on outcomes, reviewers may not be able to combine trial results in a meta-analysis. A review of 127 RCTs of psychosocial interventions for alcohol misuse and another review of 17 RCTs of mental health treatments for juveniles demonstrated deficiencies in reporting randomisation sequence generation, blinding, distinctions of primary and secondary

outcomes, and sample size calculations.<sup>89,90</sup> Another review of educational reports found no studies that adequately reported allocation concealment,<sup>87</sup> and reports in criminology often lack information about randomisation procedures.<sup>82</sup> Considering that many social and psychological interventions have relatively modest effect sizes, inadequate reporting of these factors related to internal validity frustrates the critical appraisal of their rationale and methodology.

Limitations in reporting social and psychological intervention trials also relate to external validity. A systematic review of intervention evaluations in social work revealed that only 3% of the 63 included studies gave sufficient detail to determine how the intervention protocol was actually implemented.<sup>66</sup> In a review of 28 RCTs in clinical psychology, no study provided data on adverse events.<sup>75</sup> Measuring adverse events is particularly important to the principle “do no harm” espoused by researchers and practitioners in this area.<sup>11</sup> An examination of 99 outcome studies investigating public health interventions for HIV risk reduction found that 43% of all studies did not report the theoretical basis of the intervention, despite the importance of using theory to develop effective interventions. Moreover, only 33% of studies reported data on the total number of intervention sessions attended by participants (e.g., 8 of 10 sessions), 70% of studies reported the total number of hours spent in treatment by participants (e.g., 16 hours out of a total 20), and 41% reported the duration for intervention delivery (e.g., 2 months of weekly treatment).<sup>88</sup> Such limited detail reduces the ability of practitioners to replicate these interventions and assess whether they are appropriate for their professional contexts.

These reviews collectively suggest that the reporting quality of social and psychological intervention RCTs could be improved. Until reporting inadequacies are resolved, practitioners, policy makers, and researchers may advocate interventions with unreliable evidence, as low-quality reports of RCTs tend to provide exaggerated effect

estimates of 35% on average.<sup>91</sup> Biased RCT and review results may invalidate subsequent research agendas, practice guidelines, and policy.<sup>92</sup> Vulnerable populations may suffer unnecessarily because providers and policy-makers have insufficient or inaccurate information from researchers on relevant studies to inform decision-making.<sup>93</sup> Accurately reporting the essential details of an interventional trial is thus a scientific and ethical responsibility.<sup>94</sup> Numerous reviews<sup>79</sup> demonstrating recent successful efforts to improve the reporting of medical RCTs proffer a solution: high quality reporting guidelines.

### ***5.3 Reporting Guidelines for Health Research***

For nearly 20 years, medical researchers have made a concerted effort to improve the reporting of health research by developing explicit standards for transparently describing the essential features of research studies. These reporting standards are not requirements that determine the format for writing or designing a study, but rather are recommendations on the content that study reports should consistently and transparently include to allow for critical appraisal.<sup>59</sup> In other words, standards do not prescribe research conduct, but suggest the *items of information* that are necessary to understand how a study was conducted. While researchers may be aware of the important components of a study, reporting standards serve as a safeguard to prevent seemingly mundane yet extremely important information from being overlooked or forgotten.<sup>95</sup> Adherence to these standards helps to ensure that research can effectively inform practice and policy with interventions that are accountable to robust evidence.

Reporting standards are generally collated and disseminated through reporting guidelines. A reporting guideline is “a checklist, flow diagram, or explicit text to guide authors in reporting a specific type of research, developed using explicit methodology.”<sup>96</sup> These guidelines provide authors with standards and recommendations on what to include

in an empirical research article. Guidelines usually target a minimum set of study details that: (a) are related to important research biases, (b) are often poorly reported, and/or (c) stakeholders agree should be reported. Reporting guidelines thus aim to produce research publications that have sufficient detail to interpret and apply study results.

The most well-known, influential, and widely used reporting guideline for RCTs is the Consolidated Standards of Reporting Trials (CONSORT) Statement. The CONSORT Statement is an evidence-based guideline that assists authors in clearly and sufficiently reporting essential information about all trials, though it particularly focuses on individually randomised, two-group parallel RCTs.<sup>91,96</sup> The current, revised version of the CONSORT Statement—the CONSORT 2010 Statement—provides a checklist of 25 reporting standards (see Figure 3), a diagram for documenting the flow of participants through a trial (see Figure 4), and examples of desired reporting.<sup>97</sup> These reporting standards reflect study details that should be included in an RCT to enable trial replication, assess threats to internal and external validity, extract data from the published report for reviews, and meet ethical obligations to maximise research utility.<sup>76</sup> Each checklist item was selected either because of empirical evidence indicating that *not* reporting the associated information is related to biased treatment effect estimates, and/or because the information is critical for assessing the reliability or relevance of trial findings. The flow diagram is intended to depict the number of participants through the stages of a trial: enrolment, allocation, follow-up, and data analysis. By providing this minimal list of important standards, the CONSORT Statement offers a consistent method for authors to prepare high quality publications of trial findings, facilitating their accurate, complete, and transparent reporting. It also assists peer-reviewers, editors, funding bodies, and other consumers of research in critically appraising and interpreting RCT publications and protocols.<sup>96</sup>

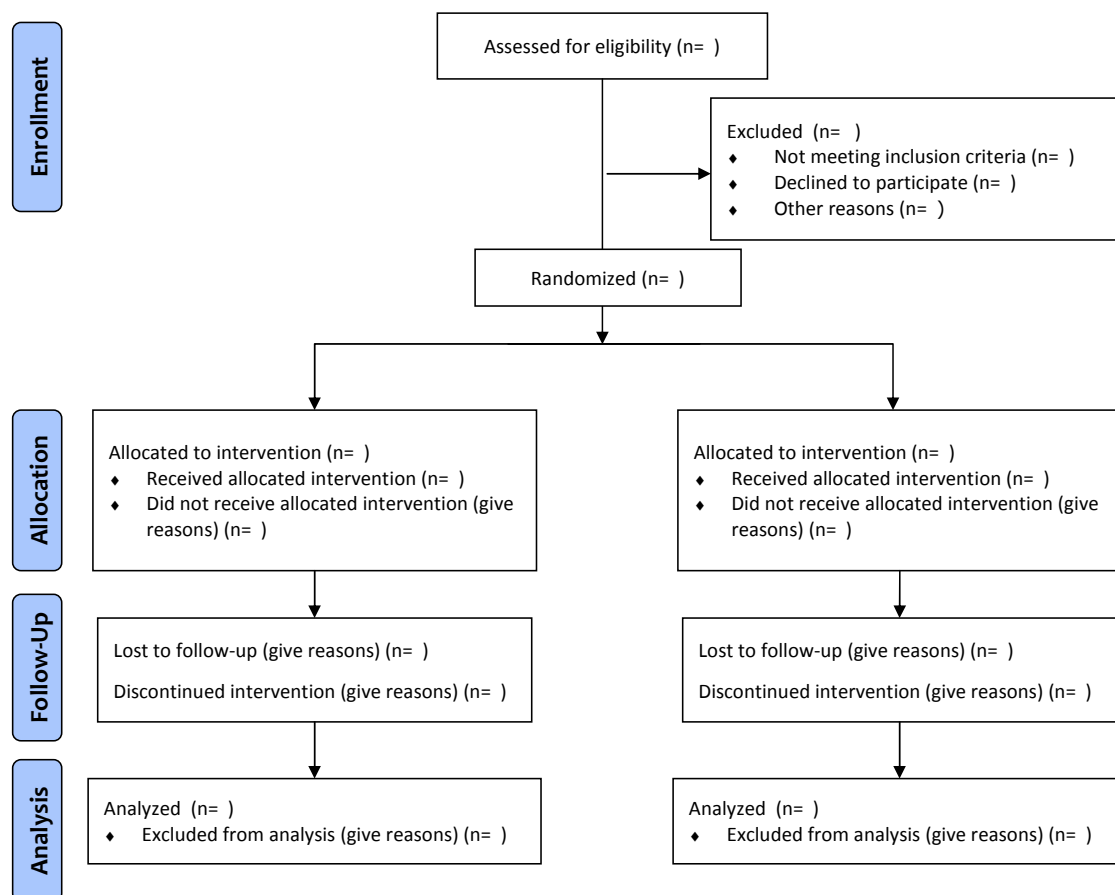
**Figure 3: CONSORT 2010 Statement Checklist**<sup>97</sup>

Section/Topic	Item	Checklist item	Page
<b>Title and abstract</b>			
	1a	Identification as a randomised trial in the title	_____
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	_____
<b>Introduction</b>			
Background and objectives	2a	Scientific background and explanation of rationale	_____
	2b	Specific objectives or hypotheses	_____
<b>Methods</b>			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	_____
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	_____
Participants	4a	Eligibility criteria for participants	_____
	4b	Settings and locations where the data were collected	_____
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	_____
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	_____
	6b	Any changes to trial outcomes after the trial commenced, with reasons	_____
Sample size	7a	How sample size was determined	_____
	7b	When applicable, explanation of any interim analyses and stopping guidelines	_____
Randomisation: Sequence generation	8a	Method used to generate the random allocation sequence	_____
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	_____
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	_____
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	_____

Section/Topic	Item	Checklist item	Page
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	_____
	11b	If relevant, description of the similarity of interventions	_____
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	_____
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	_____
<b>Results</b>			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	_____
	13b	For each group, losses and exclusions after randomisation, together with reasons	_____
Recruitment	14a	Dates defining the periods of recruitment and follow-up	_____
	14b	Why the trial ended or was stopped	_____
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	_____
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	_____
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	_____
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	_____
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	_____
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	_____
<b>Discussion</b>			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	_____
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	_____
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	_____

Section/Topic	Item	Checklist item	Page
<b>Other information</b>			
Registration	23	Registration number and name of trial registry	_____
Protocol	24	Where the full trial protocol can be accessed, if available	_____
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	_____

**Figure 4: CONSORT 2010 Statement Flowchart<sup>97</sup>**



Since its original publication in 1996, the CONSORT Statement has increased awareness of issues surrounding proper reporting of research, setting a higher industry standard. Evidence is accumulating that the CONSORT Statement has improved the reporting and utility of thousands of RCTs, principally in medicine.<sup>96</sup> An early evaluation of the first CONSORT Statement<sup>98</sup> showed that reporting in the *BMJ*, *Lancet*, and *JAMA* significantly improved after the publication of CONSORT.<sup>91</sup> Later systematic reviews

revealed that articles in medical journals instructing authors to use CONSORT when reporting RCTs are significantly better at reporting the method of sequence generation, allocation concealment, participant flow, and overall number of reporting standards than articles in journals that do not refer to CONSORT.<sup>99,100</sup> These effects remain even after controlling for the impact factor of journals and study outcome.<sup>101</sup> When utilised, the CONSORT Statement appears to be achieving the originally intended improvement of the reporting of RCTs. In response to these findings, several hundred biomedical journals and several international editorial groups, including the World Association of Medical Editors and the Council of Science Editors, have endorsed the CONSORT Statement.<sup>102</sup>

#### ***5.4 A CONSORT Extension for Social and Psychological Intervention RCTs?***

Although over 600 journals now request or recommend the use of CONSORT, this guideline must be implemented much more widely and routinely in order to reach its full potential. One important way to ensure this is to extend or augment CONSORT to specific interventions or trial designs.<sup>103</sup> Trial design extensions include versions of the CONSORT Statement for cluster trials<sup>55</sup> and pragmatic trials.<sup>104</sup> Current intervention extensions of the Consolidated Standards of Reporting Trials (CONSORT) Statement distinguish themselves by the techniques involved in the intervention. Compared to pharmacological techniques or drugs,<sup>97</sup> these intervention extensions are specifically for non-pharmacologic medical treatments (e.g., surgery, medical devices),<sup>105</sup> herbal medicinal products,<sup>106</sup> and acupuncture treatments,<sup>107</sup> all of which primarily target mechanisms emphasised in various areas of medical science, both traditional and “alternative and complementary”.

The CONSORT Extension for Non-Pharmacologic Treatments (NPT)<sup>105,108</sup> warrants specific discussion, given both its high-quality and its pertinence to this thesis project. By definition, social and psychological interventions are “non-pharmacologic

treatments”, in that they do not involve the use of drugs to influence mechanisms of interest. However, based on the disciplinary expertise of those involved in developing CONSORT-NPT, the examples provided in published guidance, and the journals of publication, social and psychological interventions as defined in this thesis differ considerably from those targeted in CONSORT-NPT by the mechanisms being targeted. Namely, CONSORT-NPT (and CONSORT in general) has a particular focus on medical treatments that work through biological mechanisms as opposed to social and psychological mechanisms. The dependency of active ingredients to work through mechanisms inextricable with human agency is the key differentiating factor. Consequently, from a disciplinary point of view, NPT has more relevance to those trained and working in biomedical sciences, whereas an extension for social and psychological interventions would appeal to social and behavioural scientists. This difference in focus results in quite significant contrasts in the terminology, theories, concepts, intervention techniques, contexts, and methods involved in interventions targeted by CONSORT-NPT versus those interventions targeted by this project. See Table 1 for a list of social and psychological intervention examples in the Campbell Collaboration library.

Many aspects of RCTs evaluating social and psychological interventions are shared with non-pharmacological medical treatment RCTs rather than differing from them. It was fully appreciated at the outset of this project that the important aspects that *do* differ in social and psychological intervention trials may not necessarily be sufficient to justify a new reporting guideline rather than simply seeking to better disseminate CONSORT-NPT to social and behavioural scientists. Moreover, researchers in biomedical sciences sometimes incorporate social and psychological elements to interventions, such as physiotherapists who require client compliance with prescribed exercises, doctors who require patient adherence to drug schedules, or even psychiatrists who combine

**Table 1. Social and Psychological Interventions from the Campbell Library<sup>109</sup>**

<b>Interventions</b>	<b>Theories</b>	<b>Techniques</b>	<b>Settings</b>	<b>Mechanisms</b>	<b>Outcomes</b>	<b>Disciplines</b>
Active Labour Market Programmes	Equilibrium Unemployment Theory	Job search assistance	Social services departments	Job search efficiency	Exit from unemployment	Economics
Brief Motivational Interventions	Transtheoretical Model of Change	Motivational interviewing	Online	Motivation, knowledge, skills	Alcohol and drug use	Psychology
Cash-Based Approaches	Paternalism Theory	Cash-transfers	Humanitarian crises in LMICs	Increased purchasing power	Household income	Public Health
Gang Resistance Programmes	Self-Control Theory	Cognitive-behavioural training, education in conflict resolution	High schools	Social skills	Gang membership	Criminology
Mentoring Interventions	Social Learning Theory	Interactive relationship between mentor and mentee	Community organisations	Self-efficacy, societal bonds	Juvenile delinquency	Criminology
Montessori Education	Constructivism	Student-chosen school activities	Elementary schools	Experiential knowledge	Academic achievement	Education
"No Excuses" Charter Schools	Laissez-Faire Education	Rigorous, strict structure	Inner-city, urban areas	Knowledge and character	Literacy and numeracy	Education
Preventive Home Visits	Health Promotion Model	Multidimensional geriatric assessment	Homes in the community	Psychosocial and environmental risk factors	Institutionalisation	Public Health
Psychoanalytic psychotherapy	Psychoanalytic Theory	Talk therapy	Therapist offices	Insight on unconscious difficulties	Post traumatic stress disorder (PTSD)	Psychology
Skills-Based Intimate Partner Violence Prevention	Feminist Theories of Gender Inequality	Modelling positive relationship skills	Universities	Communication skills, self-esteem	Dating violence	Social Work

pharmacotherapies with psychotherapies. Conversely, some psychological researchers conduct experiments looking to alter underlying neurobiology, for example via interventions using neuro-feedback techniques.

The distinction then, between biology on the one hand, and psychology and sociology on the other, entails a social construction, as it involves splitting a reality that is actually connected so as to make its constituent components more comprehensible. For example, psychological processes are emergent phenomena of biology (e.g., neural systems), social interactions are emergent of psychological processes (e.g., individual behaviours), and intervening on processes "higher up" the ontological chain impacts processes on the "lower" foundational levels (e.g., changes in cognitions alter neurochemical activity). Nonetheless, a significant proportion of intervention researchers investigate social and psychological phenomena as emergent units of reality in their own right.

These distinctions can be seen in characteristics of the developers and focus of current CONSORT guidance. The original CONSORT Statement aimed to improve the transparency of pharmacological trials, and both it and the CONSORT-NPT Extension were mainly developed by biostatisticians and biomedical researchers.<sup>69,105</sup> Few social or behavioural scientists participated in their development. While the first CONSORT Statement was published in 1996, and editors in social and psychological intervention disciplines are aware of it, efforts to promote CONSORT amongst these disciplines have largely failed. Of the journals that endorse CONSORT, only 50 are in disciplines that have a primary focus on social and psychological interventions. Whilst CONSORT has been endorsed by the Institute of Educational Sciences<sup>82</sup> and has been used to inform reporting guidelines for the American Psychological Association,<sup>69</sup> many social and behavioural scientists believe it needs amending for their research.<sup>65,75</sup>

Many researchers have proposed that social and psychological intervention trials may benefit from tailored CONSORT reporting standards specific to their areas.<sup>12,65,87</sup> In addition, members of previous CONSORT groups, as well as journal editors and researchers in the social and behavioural sciences, have indicated support for including social and psychological intervention stakeholders in developing a new, targeted guideline to, among other things, improve “buy-in” and uptake in these disciplines.<sup>87,110</sup> Furthermore, several reviews identify the need for a new reporting guideline for social and psychological intervention trials,<sup>65,66,82,87,88</sup> arguing that social and psychological interventions represent another unique intervention category that is not adequately covered by existing CONSORT guidelines.<sup>111</sup>

## 6. Conclusion

A project is needed to explore the possibility of a new reporting guideline that is sensitive to the concepts and terminology more familiar and acceptable to social and behavioural scientists conducting intervention research. Namely, to improve the reporting of social and psychological intervention RCTs, a new CONSORT extension for these trials may be a solution. The impact of CONSORT and the recent proliferation of publications on reporting quality in social and behavioural sciences indicates that such a guideline would be well-received by various stakeholders in social and psychological intervention research fields.

It is important to note that RCTs are not the only valid method for evaluating interventions,<sup>40,112,113</sup> nor are they the only method that can benefit from better reporting.<sup>114</sup> The importance of reporting standards for other types of research has been documented, including standards for observational studies,<sup>115</sup> quasi-experimental studies,<sup>116</sup> qualitative studies,<sup>117,118</sup> and methods from other paradigms that evaluate social and psychological

interventions.<sup>119,120</sup> Nonetheless, RCTs are increasingly used to evaluate social and psychological interventions, and results of these RCTs and of systematic reviews synthesising these RCTs are increasingly informing practice and policy decision-making. As such, RCTs are a strategic starting point for improving reports of many types of studies in this area, particularly if RCTs are to remain the “gold standard” of the intervention research community.<sup>110</sup> The rigour and widespread use of CONSORT could make an extension for social and psychological interventions one apt solution to address the gap in reporting standards and reporting quality for these trials.

## References

1. Grant S, Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo-Wilson E. Developing a reporting guideline for social and psychological intervention trials. *Research on Social Work Practice*. 2013;23(6):595-602.
2. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: Suggested framework and systematic review. *BMJ*. 2006;333:346-349.
3. Oakley A, Strange V, Bonell C, Allen E, Stephenson J. Process evaluation in randomised controlled trials of complex interventions. *BMJ*. 2006;332:413-416.
4. Brady HE. Causation and explanation in social science. In: Coolier D, Brady HE, Box-Steffensmeier JM, eds. *The Oxford handbook of political methodology*. Oxford: Oxford University Press; 2008.
5. Munro A, Bloor M. Process evaluation: The new miracle ingredient in public health research? *Qualitative Research*. 2010;10(6):699-713.
6. Bowling A. *Research methods in health: Investigating health and health services*. New York: McGraw-Hill; 2009.
7. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth; 2002.
8. Chalmers I. Fisher and Bradford Hill: Theory and pragmatism? *International Journal of Epidemiology*. 2003;32:922-924.
9. Bradford Hill A. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*. 1965;1:295-300.
10. Höfler M. The Bradford Hill considerations on causality: A counterfactual perspective. *Emerging Themes in Epidemiology*. 2005;2:11.
11. Chalmers I. Trying to do more good than harm in policy and practice: The role of

- rigorous, transparent, up-to-date evaluations. *The ANNALS of the American Academy of Political and Social Science*. 2003;589(22):22-40.
12. Boutron I, John P, Torgerson DJ. Reporting methodological items in randomized experiments in political science. *The ANNALS of the American Academy of Political and Social Science*. 2010;628(112):112-131.
  13. Tucker-Drob EM. Individual differences methods for randomized experiments. *Psychological Methods*. 2011;16(3):298-318.
  14. Hayward RA, Kent DM, Vijan S, Hofer TP. Reporting clinical trial results to inform providers, payers, and consumers. *Health Affairs*. 2005;24(6):1571-1581.
  15. Moher D, Hopewell S, Schultz KF, Altman DG. Resources for authors of reports of randomized trials: Harnessing the wisdom of authors, editors, and readers. *Trials*. 2011;12:98.
  16. Cochrane AL. *Effectiveness and efficiency: Random reflections on health services*. Nuffield Provincial Hospitals Trust: The Rock Carling Fellowship; 1972.
  17. Ashcroft RE. Current epistemological problems in evidence based medicine. *Journal of Medical Ethics*. 2004;30:131-135.
  18. Sackett DL, Rosenberg WMC, Muir Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. *BMJ*. 1996;312(7023):71-72.
  19. Sehon SR, Stanley DE. A philosophical analysis of the evidence-based medicine debate. *BMC Health Services Research*. 2003;3:14.
  20. Howick J. *The philosophy of evidence-based medicine*. Chichester: Wiley-Blackwell; 2011.
  21. Macintyre S, Petticrew M. Good intentions and received wisdom are not enough: The need for controlled trials in public health. *Journal of Epidemiology & Community Health*. 2011;54:802-803.

22. Akobeng AK. Understanding randomised controlled trials. *Archives of Disease in Childhood*. 2005;90:840-844.
23. Gibbs L, Gambrill E. Evidence-based practice: Counterarguments to objections. *Research on Social Work Practice*. 2002;12:452-476.
24. Torgerson DJ, Torgerson CJ. *Designing randomised trials in health, education and the social sciences: An introduction*. Palgrave Macmillan; 2008.
25. Bloom HS. *Learning more from social experiments: Evolving analytic approaches*. Russel Sage Foundation; 2005.
26. Gambrill E. Evidence-based practice: Sea change or emperor's new clothes? *Journal of Social Work Education*. 2003;39:3-23.
27. Akobeng AK. Principles of evidence based medicine. *Archives of Disease in Childhood*. 2005;90:837-840.
28. Macdonald G, Sheldon B. Contemporary studies of the effectiveness of social work. *British Journal of Social Work*. 1992;22(6):615-643.
29. Petrosino A, Turpin-Petrosino C, Buehler J. "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database of Systematic Reviews*. 2009;1:CD002796.
30. Oakley A. Resistances to 'new' technologies of evaluation: Education research in the UK as a case study. *Evidence & Policy*. 2006;2(1):63-87.
31. APA Presidential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. *American Psychologist*. 2006;61:271-283.
32. Substance Abuse and Mental Health Services Administration. *SAMHSA's national registry of evidence-based programs and practices*. Retrieved 3 June 2014 from <http://www.nrepp.samhsa.gov/>.
33. Institute of Education Sciences. *IES's what works clearinghouse*. Retrieved 3 June

- 2014 from <http://ies.ed.gov/ncee/wwc/default.aspx>.
34. Chalmers I. If evidence-informed policy works in practice, does it matter if it doesn't work in theory? *Evidence & Policy*. 2005;1:227-242.
  35. Gambrill E. Evidence-based practice and policy: Choices ahead. *Research on Social Work Practice*. 2006;16(3):338-357.
  36. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: How to practice and teach EBP (2nd ed.)*. New York: Churchill Livingstone; 2000.
  37. Bonell C. The utility of randomized controlled trials of social interventions: An examination of two trials of HIV prevention. *Critical Public Health*. 2002;12:321-334.
  38. Melia S. *Do randomised control trials offer a solution to 'low quality' transport research?* [Working paper]. University of the West of England, Bristol. 2014.
  39. Macintyre S, & Petticrew, M. Good intentions and received wisdom are not enough. *Journal of Epidemiology & Community Health*. 2000;54:802-803.
  40. Hammersley M. Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice. *Evidence & Policy*. 2005;1(1):85-100.
  41. Macdonald G. Evidence-based social care: Wheels off the runway? *Public Money & Management*. 1999;19(1):25-32.
  42. Campbell DT. Assessing the impact of planned social change. *Evaluation and Program Planning*. 1979;2:67-90.
  43. Bonell CP, Hargreaves J, Cousens S, et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. *Journal of Epidemiological Community Health*. 2011;65:582-587.
  44. May C. Towards a general theory of implementation. *Implementation Science*.

- 2013;8:18.
45. Weiss MJ, Bloom HS, Brock T. *A conceptual framework for studying the sources of variation in program effects*. New York, NY: MDRC; 2013.
  46. Pawson R, Greenhalgh T, Harvey G, Walshe K. *Realist synthesis: An introduction*. University of Manchester: ESRC Research Methods Programme; 2004.
  47. Medical Research Council. *Developing and evaluating complex interventions: New guidance*. London: Medical Research Council; 2008.
  48. Walker R, Hoggart L, Hamilton G. Observing the implementation of a social experiment. *Evidence & Policy*. 2008;4(3):183-203.
  49. Shepperd S, Lewin S, Straus S, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Medicine*. 2009;6(8):31000086.
  50. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008;337:979-983.
  51. Medical Research Council. *A framework for development and evaluation of RCTs for complex interventions to improve health*. London: Medical Research Council; 2000.
  52. Moore G, Audrey S, Barker M, et al. Process evaluation in complex public health intervention studies: The need for guidance. *Journal of Epidemiology & Community Health*. 2013;68(2):101-102.
  53. Flay BR, Biglan A, Boruch RF, et al. Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*. 2005;6:151-175.
  54. Rycroft-Malone J, Burton CR. Is it time for standards for reporting on research about implementation? *Worldviews on Evidence-Based Nursing*. 2011;8(4):189-190.
  55. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: Extension to cluster randomised trials. *BMJ*. 2004;328:702-708.

56. Fraser MW, Galinsky MJ, Richman JM, Day SH. *Intervention research: Developing social programs*. New York, NY: Oxford University Press; 2009.
57. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: A major source of bias in randomised controlled trials of treatments for schizophrenia. *British Journal of Psychiatry*. 2000;176:249-252.
58. Jun M, Zoungas S, Perkovic V, Webster AC. How to read a report of a randomized controlled trial. *Nephrology*. 2010;15:153-157.
59. American Educational Research Association. Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*. 2006;35:33-40.
60. Moore L, Moore GF. Public health evaluation: Which designs work, for whom and under what circumstances? *Journal of Epidemiology & Community Health*. 2011;65:596-597.
61. Cartwright N, Munro E. The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*. 2010;16:260-266.
62. Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*. 2004;23(5):443-451.
63. Hawe P, Shiell A, Riley T. Complex interventions: How "out of control" can a randomised controlled trial be? *BMJ*. 2004;328:1561-1563.
64. Hardeman W, Michie S, Fanshawe T, Prevost AT, McLoughlin K, Kinmonth AL. Fidelity of delivery of a physical activity intervention: Predictors and consequences. *Psychology & Health*. 2008;23:11-24.
65. McGrath PJ, Stinson J, Davidson K. Commentary: The Journal of Pediatric Psychology should adopt the CONSORT statement as a way of improving the evidence base in pediatric psychology. *Journal of Pediatric Psychology*. 2003;28:169-171.

66. Naleppa MJ, Cagle JG. Treatment fidelity in social work intervention research: A review of published studies. *Research on Social Work Practice*. 2010;20:674-681.
67. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ*. 2008;336:1472-1474.
68. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: Effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*. 2012;13(6):CD000259.
69. APA JARS Group. Reporting standards for research in psychology: Why do we need them? what might they be? *American Psychologist*. 2008;63:839-851.
70. Evans T, Brown H. Road traffic crashes: Operationalizing equity in the context of health sector reform. *Injury Control and Safety Promotion*. 2003;10(1-2):11-12.
71. Moore L. Research design for the rigorous evaluation of complex educational interventions: Lessons from health services research. *Building Research Capacity*. 2002;1:4-5.
72. Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*. 2012;75(12):2299-2306.
73. Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F. *Implementation research: A synthesis of the literature*. Tampa, FL: National Implementation Research Network, University of South Florida; 2005.
74. Wang S, Moss JR, Hiller JE. Applicability and transferability of interventions in evidence-based public health. *Health Promotion International*. 2006;21(1):76-83.
75. Stinson JN, McGrath PJ, Yamada JT. Clinical trials in the Journal of Pediatric Psychology: Applying the CONSORT statement. *Journal of Pediatric Psychology*. 2003;28:159-167.

76. Davidson KW, Goldstein M, Kaplan RM, et al. Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine*. 2003;26:161-171.
77. Tugwell P, Petticrew M, Kristjansson E, et al. Assessing equity in systematic reviews: Realising the recommendations of the Commission on Social Determinants of Health. *BMJ*. 2010;341:c4739.
78. Welch V, Petticrew M, Tugwell P, et al. PRISMA-Equity 2012 extension: Reporting guidelines for systematic reviews with a focus on health equity. *PLoS Medicine*. 2012;9(10):e1001333.
79. Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *Journal of Clinical Epidemiology*. 2011;64:136-144.
80. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology*. 2004;4:22.
81. Glasziou P, Chalmers I, Altman DG, et al. Taking healthcare interventions from trial to practice. *BMJ*. 2010;341:384-387.
82. Perry AE, Weisburd D, Hewitt C. Are criminologists describing randomized controlled trials in ways that allow us to assess them? Findings from a sample of crime and justice trials. *Journal of Experimental Criminology*. 2010;6:245-262.
83. Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR Network's Survey of Guideline Authors. *PLoS Medicine*. 2008;5:869-874.
84. MST Services. Accessed at <http://mstservices.com/> on 1 August 2011.
85. Littell JH, Campbell M, Green S, Toews B. Multisystemic therapy for social,

- emotional, and behavioral problems in youth aged 10-17. *Cochrane Database of Systematic Reviews*. 2005;19(4):CD004797.
86. Higgins J, Green S, eds. *Cochrane handbook for systematic reviews of interventions* Version 5.1.0 5.1.0 [updated March 2011]. Chichester, UK: Wiley-Blackwell; 2011.
87. Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. A comparison of RCTs in health and education. *British Educational Research Journal*. 2005;31(6):761–785.
88. Semaan S, Kay L, Strouse D, et al. A profile of U.S.-based trials of behavioral and social interventions for HIV risk reduction. *Journal of Acquired Immune Deficiency Syndromes*. 2002;30:S30-S50.
89. Ladd BO, McCrady BS, Manuel JK, Campbell W. Improving the quality of reporting alcohol outcome studies: Effects of the CONSORT statement. *Addictive Behaviors*. 2010;35:660-666.
90. Perry AE, Johnson M. Applying the Consolidated Standards of Reporting Trials (CONSORT) to studies of mental health provision for juvenile offenders: A research note. *Journal of Experimental Criminology*. 2008;4:165-185.
91. Moher D, Jones A, Lepage L, for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA*. 2001;285:1992-1995.
92. Moher D. Reporting research results: A moral obligation for all researchers. *Canadian Journal of Anesthesiology*. 2007;54:331-335.
93. Chalmers I. Meeting the research information needs of patients and clinicians more effectively. *Achieving Transparency in Reporting Health Research 2008*.
94. Ghersi D. Research funders and research reporting. *Achieving Transparency in Reporting Health Research 2008*.
95. Gawande A. *The checklist manifesto: How to get things right*. London: Profile Books;

- 2010.
96. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
97. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698-702.
98. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637-639.
99. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *MJA*. 2006;185:263-267.
100. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews*. 2012;11:MR000030.
101. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Controlled Clinical Trials*. 2002;23:380-388.
102. Moher D, Altman DG, Schulz KF, Elbourne DR. Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *CMAJ*. 2004;171(4):349-350.
103. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: Reporting

- guidelines and the EQUATOR Network. *BMC Medicine*. 2010;8:24.
104. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ*. 2008 337:a2390.
105. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT Group. Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: Explanation and Elaboration. *Annals of Internal Medicine*. 2008;148:295-309.
106. Gagnier J, Boon H, Rochon P, Barnes J, Moher D, Bombardier C. Reporting random controlled trials of herbal medicines. *Explore: The Journal of Science and Healing*. 2006;2(2):143-149.
107. MacPherson H, White A, Cummings M, et al. Standards for Reporting Interventions in Controlled Trials of Acupuncture: The STRICTA recommendations. *Journal of Alternative & Complementary Medicine*. 2002;8(1):85-89.
108. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
109. Campbell Collaboration. Accessed at [campbellcollaboration.org](http://campbellcollaboration.org) on 1 June 2014.
110. Gill CE. Missing links: How descriptive validity impacts the policy relevance of randomized controlled trials in criminology. *Journal of Experimental Criminology*. 2011;7(3):201-224.
111. Mayo-Wilson E. Reporting implementation in randomized trials: Proposed additions to the Consolidated Standards of Reporting Trials statement. *American Journal of Public Health*. 2007;97:630-633.
112. Pawson R, Wong G, Owen L. Known knowns, known unknowns, unknown

- unknowns: The predicament of evidence-based policy. *American Journal of Evaluation*. 2011;32:518-546.
113. Webb SA. Some considerations on the validity of evidence-based practice in social work. *British Journal of Social Work*. 2001;31:57-79.
114. Goldbeck L, Vitiello B. Reporting clinical trials of psychosocial interventions in child and adolescent psychiatry and mental health. *Child and Adolescent Psychiatry and Mental Health*. 2011;5:4.
115. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*. 2007;147(8):573-577.
116. Des Jarlais DC, Lyles C, Crepaz N, the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*. 2004;94(3):361-366.
117. Elliott R. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*. 1999;38:215-229.
118. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007;19(6):349-357.
119. Greenhalgh T, Wong G, Westhorp G, Pawson R. Protocol--Realist and meta-narrative evidence synthesis: Evolving standards (RAMESES). *BMC Medical Research Methodology*. 2011;11:115.
120. Stern E, Stame N, Mayne J, Forss K, Davies R, Befani B. *Broadening the range of designs and methods for impact evaluation*. London: Department for International Development; 2012.

## Chapter 2: Protocol for Developing the CONSORT-SPI Extension

*NB: This chapter has been published in an alternative form in Implementation Science<sup>1</sup>*

### 1. Introduction

The previous chapter provided an overview of developments in social and psychological intervention research, making the conceptual case that a new reporting guideline may be needed for RCTs in this area. A new CONSORT extension that draws on previous reporting guidance, up-to-date scientific literature, and stakeholder involvement and insight, has the potential to significantly improve the reporting of social and psychological intervention RCTs.

The purpose of this second chapter is to outline the methods for this project overall, delineating the corresponding activities within the project plan that will constitute the work of and chapters for this thesis. The proposed methods follow recommended techniques for developing and disseminating reporting guidelines, which are based on a survey of developers of 37 different reporting guidelines.<sup>2,3</sup> Aspects of these best practices have been previously used to develop the CONSORT Statement<sup>4</sup> and some of its extensions,<sup>5-7</sup> the SPIRIT Statement for trial protocols,<sup>8</sup> and several other reporting guidelines.<sup>9,10</sup> Ethical approval of the proposed project methods has been granted by the Departmental Research Ethics Committee (DREC) at the Department of Social Policy and Intervention, University of Oxford (Ref: 2011-12\_83). As this chapter represents the project protocol drafted by the DPhil candidate prior to the project launch, this chapter has been kept in “*future tense*”, except where the specific roles of the DPhil candidate are noted.

#### 1.1 Objective

The objective of this project is to develop an extension of the CONSORT Statement

for Social and Psychological Interventions: CONSORT-SPI. This new guideline will include a checklist of reporting items, a participant flowchart, and an Explanation and Elaboration (E&E) document, all of which combined will offer authors recommendations to accurately, comprehensively, and transparently describe randomised controlled trials (RCTs) of social and psychological interventions.

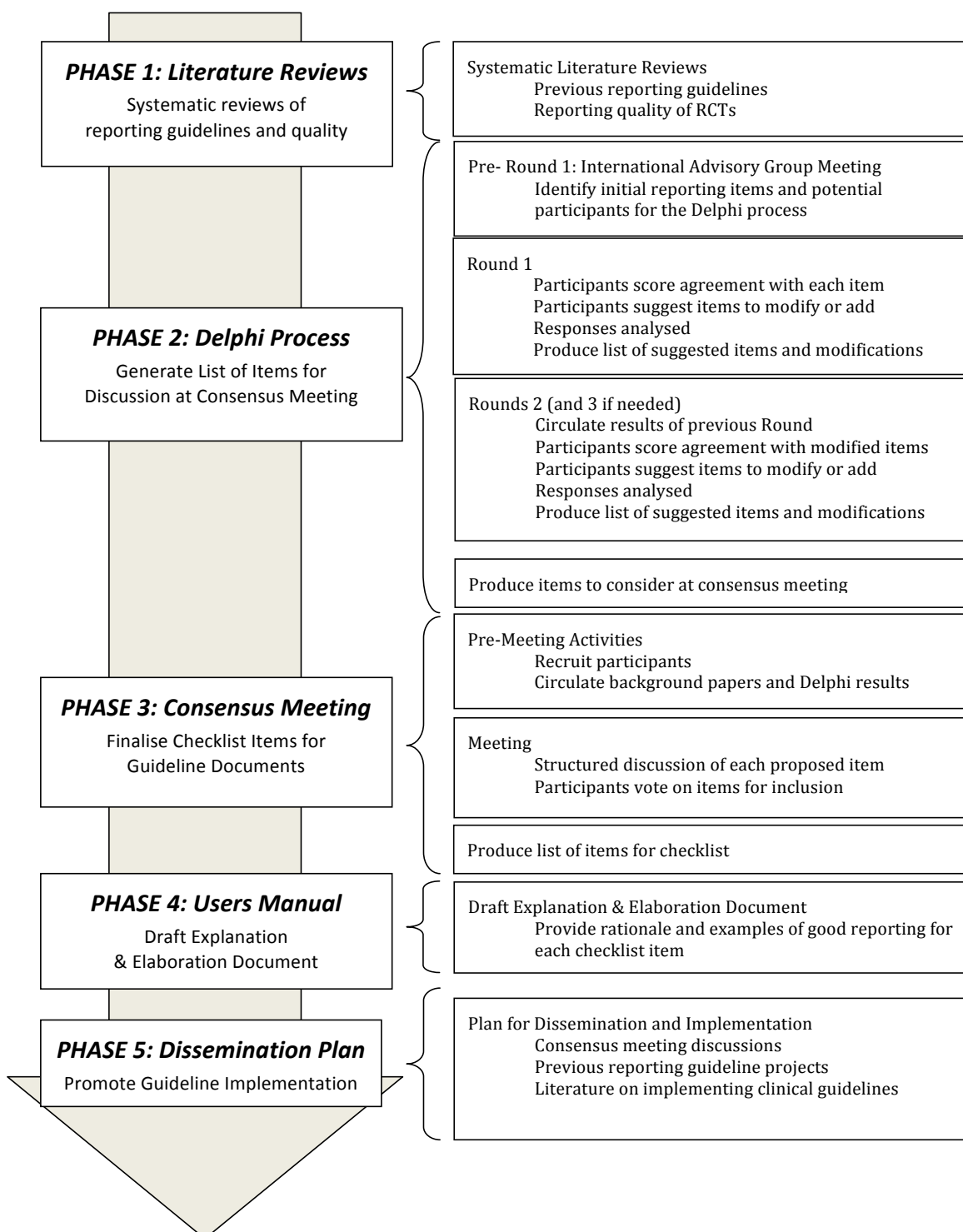
The following reflect the primary research questions within each project phase as they correspond to chapters of this DPhil thesis:

- Phase 1: is there empirical evidence to support the need for a CONSORT extension for social and psychological interventions?
- Phase 2: what reporting standards should be considered for the CONSORT-SPI guideline documents?
- Phase 3: what items should be included in the CONSORT-SPI checklist, and what is the remit of this guideline?
- Phase 4: what is the importance of included CONSORT-SPI checklist items, and how do authors adhere to them when reporting social and psychological intervention RCTs?
- Phase 5: how should CONSORT-SPI be disseminated and implemented for maximum impact?

## ***1.2 Project Overview and Thesis Outline***

This project involves five phases: systematic literature reviews, a Delphi process, a consensus development conference, writing up the guideline documents, and guideline dissemination and implementation (see Figure 1). The DPhil thesis will consist of key activities at each phase of the project plan, corresponding to the primary research questions in the previous section.

**Figure 1. Workflow for Thesis on CONSORT-SPI Project**



Phase 1 will involve systematic reviews on previous reporting guidelines and RCT reporting quality.<sup>11</sup> These reviews will empirically assess the need for a new reporting guideline for social and psychological intervention RCTs.

Phases 2 and 3 will involve consensus development methods to generate the content of the CONSORT-SPI Extension. Rather than relying on the views of a small group to decide reporting guideline content, formal consensus development methods are useful to provide recommendations that are not arbitrary. These methods are increasingly being employed to assist the effectiveness of decision-making in health and social care when there is contradictory opinion or insufficient information.<sup>12</sup> These processes use techniques that capture the advantages of group decision-making while overcoming biases associated with less structured group methods.<sup>13</sup> Previous research shows that these methods are appropriate and successful ways to synthesise expertise and research evidence for the purposes of this project,<sup>14</sup> and are beneficial to use in combination.<sup>15</sup>

For Phase 2, an online, international Delphi process will be conducted to *explore* the extent of stakeholder agreement about the importance of particular reporting items in social and psychological intervention trials. For Phase 3, a formal consensus meeting will take place to *develop* consensus for a minimum list of reporting standards for social and psychological intervention RCTs.<sup>14</sup>

The final Phases 4 and 5 of the project involve drafting the CONSORT-SPI checklist, the Explanation and Elaboration (E&E) Document, and the formal dissemination and implementation plan. For Phase 4, this thesis will include a full explication of the rationale for each reporting item in the CONSORT-SPI checklist; this chapter will serve as an initial template for the CONSORT-SPI E&E document. For Phase 5, this thesis will map current and proposed guideline dissemination activities onto frameworks for guideline implementation.<sup>16</sup> This discussion will form the basis of the dissemination and implementation strategy for CONSORT-SPI outputs moving forward.

### *1.3 Feasibility*

The project has received a standard grant from the Economic and Social Research Council (ESRC) to support the proposed methods. Furthermore, the success of this project depends on widespread agreement among key international stakeholders in research, policy, and practice. To ensure project success, the following were secured at the outset of the project by the DPhil candidate along with his supervisors:

- 1) a Project Executive to help guide the methods of each phase;
- 2) an International Advisory Group (IAG) of key opinion leaders across core disciplines to help coordinate and publicise the initiative (see Appendix A);
- 3) participation from high-impact journal editors for the consensus processes;
- 4) agreement from and collaboration with the CONSORT Group.

#### *1.3.1 Project Executive*

Members of the Project Executive (Paul Montgomery, Evan Mayo-Wilson, Susan Michie, Geraldine Macdonald, and David Moher, along with the DPhil candidate) are involved to various degrees in recruiting the IAG and other stakeholders, running the Delphi process and consensus meeting, writing-up the resultant guideline documents, and leading the dissemination and implementation strategy.

#### *1.3.2 International Advisory Group*

This team of opinion leaders will provide advice at each project phase, as well as help draft and disseminate the final guideline documents to their respective areas of research. Members are leading experts in social and psychological interventions across various disciplines (see Appendix A). They will help recruit stakeholders to participate in the project and identify topics to discuss at each stage. Tele-conferences with the IAG will

provide input and advice at key stages of the project, including consensus development, guideline drafting, and dissemination. The group will first meet (virtually via Skype) to nominate credible participants and potential reporting items for Round 1 of the Delphi process.<sup>12</sup> After the final round of the Delphi process, the group will be consulted regarding the draft agenda for the consensus meeting. Members of the group will be considered for the consensus meeting and will participate in drafting the guideline. They will be asked to aid in dissemination by endorsing and using the guideline, and by presenting it to relevant stakeholders in their particular disciplines.

### *1.3.3 Journal Editors*

The most widely used reporting guidelines have (a) enlisted journal editors during development and (b) acquired official journal endorsement upon completion.<sup>2</sup> Involving these key stakeholders in the development process should increase the likelihood of guideline uptake similar to the original CONSORT Statement.<sup>2,17,18</sup> To begin this effort, editors of high-impact journals in key disciplines were approached from the outset of the project, and many have agreed to participate and to potentially endorse the guideline. A list of journals that have had editors already participate or express interest in this project can be found in Appendix B.

### *1.3.4 CONSORT Group*

To increase successful uptake, new reporting guidelines for RCTs should be officially related to the CONSORT Statement. Previous reviews have found no high-impact journal that explicitly recommends an RCT reporting guideline other than the CONSORT statement.<sup>19</sup> Members of the CONSORT Group are involved in this project, and the resultant guideline will be an official extension of the CONSORT Statement. The

CONSORT Group's success, collective experience, and prominence in the reporting guideline field will help to ensure high-quality development and dissemination.

#### ***1.4 DPhil Activities in the CONSORT-SPI Project***

Each section of the following chapter will identify the specific aspects of the project phase that will be in the DPhil thesis and then provide a brief summary of the methods for DPhil activities within each project phase. These activities were specifically chosen to allow for the DPhil candidate to make a sufficient number of individual, significant contributions to the project as required for this thesis.

## **2. Project Phase 1: Systematic Reviews**

### ***2.1 Objective***

Experienced reporting guideline developers, including members of the CONSORT Group, recommend a preliminary review of relevant research literature before developing a new or extending an existing reporting guideline.<sup>2</sup> Namely, developers have three options to consider moving forward. Firstly, they can create an entirely new guideline if no previous guidance is near adequate. Secondly, they can extend previous guidance by augmenting and tailoring contents of an existing, high-quality guideline to their specific field of enquiry. Thirdly, they can simply “implement” an existing, high-quality guideline as it stands by focusing on dissemination and implementation of the guideline rather than tailoring any guideline content or developing an entirely new document.

### ***2.2 Thesis Chapter 3: Protocol for Systematic Reviews***

Chapter 3 of the thesis will constitute two systematic reviews: one on previous reporting guidelines relevant to social and psychological intervention RCTs, and another

on the reporting quality of published RCTs in widely-cited scientific journals. These reviews were led by the DPhil candidate under the guidance of his supervisors.

### *2.2.1 Review of Relevant Reporting Guidelines*

To determine the best course of action, guideline developers should systematically search for and catalogue existing, relevant reporting guidelines, and they should assess the quality of guideline development, content, and dissemination efforts.<sup>2</sup> While several researchers have called for a new CONSORT extension for social and psychological interventions,<sup>20</sup> a systematic review of the development, content, and dissemination of previous guidelines will indicate whether any single guideline is already sufficient, or whether a new guideline is truly needed. This review should identify and assess the rigour of all published reporting guidelines (not just CONSORT) that are relevant to social and psychological intervention RCTs. If an existing guideline is of sufficient quality, then it should be disseminated rather than researchers developing an entirely new guideline. If none are suitable, then the development of a new guideline or extension of a previous guideline may be the appropriate solution moving forward.

### *2.2.2 Review of RCT Reporting Quality*

Guideline developers should also collect systematic evidence on the reporting quality of the targeted research method. Though informative, previous reviews of social and psychological intervention RCT reporting quality have two primary limitations. Firstly, each individual review focused on small samples of RCTs within individual disciplines (e.g., clinical psychology, social work). A systematic review of RCTs across several disciplines is needed to clearly demonstrate deficiencies in reporting these trials overall. Secondly, each review utilised a different sub-set of known standards when

assessing reporting quality, obstructing a cohesive argument that justifies a new reporting guideline suitable for all social and psychological intervention trials. A more comprehensive systematic review would thus build upon previous investigations by (a) concurrently examining social and psychological intervention RCTs across multiple disciplines and (b) utilising a complete, unified set of published reporting standards. Such a review would have the ability to suggest that overall reporting of social and psychological intervention RCTs is inadequate and, importantly, that a single, unified set of reporting standards can be feasibly applied to these trials regardless of academic discipline.

### **3. Project Phase 2: The Delphi Process**

#### ***3.1 Objective***

The purpose of Phase 2 is to identify those areas of reporting social and psychological intervention trials that are most important for inclusion in the guideline. To involve a wide range of participants at this phase, the DPhil candidate will conduct an international Delphi process: that is, a series of sequential questionnaires answered anonymously by a panel of stakeholders with relevant expertise.<sup>21</sup> The Delphi method is a prominent group facilitation technique that involves an iterative, multistage procedure to exchange information and synthesise opinion into group consensus.<sup>12</sup> A series of structured questionnaires are completed anonymously by “expert” participants, and summarised responses from each questionnaire are provided to participants via controlled feedback after each round.

An online, modified Delphi process is preferred for Phase 2 due to its strengths in surveying areas of uncertainty and measuring consensus.<sup>14</sup> The Delphi method has several advantages compared to focus groups and questionnaires. Firstly, via anonymous responding and controlled feedback, it addresses social-cognitive biases that typically

promote conformity during group discussions. By using successive online surveys and controlled summarised feedback, participants are more likely to consider group opinions in a non-adversarial manner compared to focus group formats.<sup>12</sup> Moreover, participants in Delphi processes do not interact directly with each other, which avoids dominance by assertive or senior individuals as well as deference by passive panel members or those more junior in the group. Nonetheless, the method maintains the plurality of views generated through group discussion. Participants are still informed by the collective opinions of the group and can identify items that may have been missed or thought unimportant, allowing the opportunity for participants to change their opinions in light of feedback.<sup>22</sup> Combined, these features allow participants to consider their previous opinions, challenge received ideas, and stimulate discussion on new concepts without pressures to conform to participants of higher status.

This method is also a cost-effective way to involve a large number of international and cross-disciplinary participants.<sup>21</sup> It gathers stakeholder opinion without needing to bring participants physically together, and increases the chances of subsequent uptake of the final guideline by these stakeholders, given their participation and input at earlier stages.<sup>23</sup> Online Delphi processes have been successful in previous studies by the CONSORT Group to reduce the number of checklist items to be considered at the consensus meeting.<sup>4,10,24,25</sup>

### ***3.2 Thesis Chapter 4: The CONSORT-SPI Delphi Process Protocol***

Chapter 4 of the thesis will report on the methods and results of this project's Delphi process. The DPhil candidate took the lead in setting the agenda and conducting the IAG meeting, the recruitment of participants, and the design and analysis of the Delphi process. The *a priori* protocol for the Delphi process is detailed below.

### *3.2.1 International Advisory Group Meeting*

Before Round 1 of the Delphi process, the IAG will meet to nominate credible participants and initial items for the Round 1 questionnaire. Prior to the meeting, the group will receive the literature reviews from Phase 1 regarding (a) previous reporting guidelines and (b) the reporting quality of experiments in top social and behavioural research journals, as well as feedback from a consultation held at the 2012 Cochrane Colloquium.<sup>26</sup> Then, a finalised plan will be decided for participant selection, number of rounds, items to include in first round, and cut-offs for consensus in line with recommended techniques for guideline development<sup>2</sup> and previous Delphi processes.<sup>7,27</sup>

### *3.2.2 Selection of Participants*

To enhance credibility and widespread acceptance, the DPhil candidate will recruit informed and interested participants who represent the diverse stakeholder groups that the guideline is intended to influence.<sup>13</sup> This aims to ensure that a variety of perspectives are captured and that a multidisciplinary, international consensus will be developed, as the guideline will have an impact across many specialist fields.<sup>14</sup> The IAG will help identify an initial list of stakeholders who extensively publish, fund, or utilise this research, and a “snowball recruitment” approach will be used.<sup>4,28</sup> In order to engage those who might not be identified through snowball recruitment, the project website will also enable stakeholders to register their interest in participating. In addition, a commentary based on the DPhil candidate’s transfer of status materials and co-published in several journals invited other stakeholders to register their interest in participating.<sup>29-36</sup> Research-informed purposive sampling via these methods is intended to provide a less biased selection of participants than the Project Executive could provide alone.<sup>13</sup> Following previous Delphi processes, at least 100 participants are desired to hopefully obtain a single panel of at least

60 participants, a manageable size whilst including a diverse and significant group of stakeholders.<sup>4,10,24</sup>

Invited participants are envisioned to represent several stakeholder groups. Intervention researchers, methodologists, and guideline developers will form a substantial number of the participants. Editors of high-impact factor journals will be invited for their expertise and to ensure uptake upon completion of the guideline. Journals are most likely to endorse the guideline if leading publications champion its development and dissemination.<sup>37</sup> Funders of studies about social and psychological interventions will be invited to provide expertise and to promote use of the guideline for assessing grant applications.<sup>10</sup> A review of 61 research funders' guidelines revealed a need for more detailed guidance for trial proposals,<sup>38</sup> which should contain sufficient information for reviewers to determine the likely validity and importance of proposed research. Practitioners will help clarify issues of relevance to practice,<sup>39</sup> while policy-makers will help identify items and assist in the creation of user-friendly documents and standards. Those representing users of social and psychological interventions will help advance the relevance of research reports to the ultimate recipients of services.<sup>3,40</sup>

### *3.2.3 Structure*

Identified stakeholders will be invited to participate in a web-based Delphi process to rate the importance of suggested checklist items for the CONSORT-SPI Extension. An introductory letter will be sent to all participants. If invited participants accept, they will be provided a link to the Round 1 questionnaire, with instructions on how to complete the questionnaire, information about the deadline for completing it, and a summary of results from the literature reviews in order to ground discussion in research evidence.<sup>13</sup> The Delphi Round 1 questionnaire and other materials will only be provided to those agreeing

to participate, in order to prevent post-recruitment attrition biases.<sup>21</sup> In accordance with previous CONSORT Extensions, participants will be sent an e-mail reminder at least 48 hours prior to the completion deadline for each round of the survey; no other reminders are intended to be given.<sup>5</sup> It is estimated that up to three rounds might be needed, though the process will end prior to three rounds or continue past three rounds until *a priori* definitions of consensus are obtained or diminishing returns are evident.<sup>12</sup>

Prior to Round 1, the DPhil candidate will use the systematic reviews from Phase 1 as well as feedback from the Project Executive and IAG to generate a preliminary list of reporting items for the Delphi Round 1 questionnaire. The aim is to develop a manageable number of items while facilitating attention to those standards that might be essential for social and psychological interventions.<sup>13</sup> Based on previous guidelines, this process is expected to yield 25 to 70 checklist items for the Round 1 questionnaire.<sup>7,41</sup> Items included in the survey will be arranged under conventional section headings for scientific articles (i.e., title, abstract, methods, results, discussion, other information). Each item for consideration in the CONSORT-SPI Extension will be listed in as concise yet clear of a manner as possible to facilitate consistency of interpretation, voting, and group discussion. All participants will form a single panel. One-month limits for each round will be planned, in order to allow participants to think critically whilst maintaining project momentum.<sup>13</sup>

In Round 1, participants will use a Likert scale to rank the importance of suggested items. In a formal questionnaire survey, participants will be asked to rate, on a 1-10 scale, the necessity of including each reporting item in the CONSORT-SPI Extension. Participants will also be provided free-text comment boxes for each item so that they may provide explanations for their ratings, suggestions for modifications to the wording of items, and any additional items missing that should be considered by the Project Executive. Responses will be collated, and summarised results will be sent via email to participants

prior to Round 2 in order to enable participants to consider their initial ranking in relation to the group's responses. The median and inter-quartile range of rankings will be provided as feedback to participants due to this data format's robustness to outliers and the distribution of responses.<sup>13</sup> An analysis of the free-text comments will be performed to identify common themes and to generate additional items if suggested by participants. These themes will be used along with the quantitative results to identify which standards may be more important than others to include in the CONSORT-SPI guideline or to reassess in future rounds of the Delphi process.

Items from Round 1 that did not reach consensus will be rephrased and regrouped in the Round 2 questionnaire according to responses from Round 1.<sup>7</sup> Participants will again rank each item according to perceived importance of inclusion or exclusion in the guideline checklist. They will also be asked whether they agree with items recommended for inclusion or exclusion from Round 1, or if any essential items are still missing. Such iteration will allow participants to change their views in light of information regarding the beliefs of others in the Delphi panel. If important disagreements remain after Round 2, participants with divergent views will be asked to give additional feedback about these items, including suggestions for resolving the differences.

If a Round 3 is needed, a potential list of items for the consensus meeting will be provided to participants. Participants will be asked to rank their agreement and to provide comments. Deviant cases (e.g., those outside the inter-quartile range for an item in Round 1) will be identified and approached to discuss their responses.<sup>13</sup> Information about disagreements will be shared with the group at the consensus meeting.

At the end of the Delphi process, high-ranking items will be proposed for inclusion in the checklist during the consensus meeting. Low-ranking items will not be considered at the consensus meeting unless the Project Executive identifies issues as valuable to discuss.

Middle-ranking items will be discussed at the consensus meeting for possible inclusion or exclusion. Using these decision rules, then, a prioritised list of items for discussion at the consensus meeting will be generated.<sup>21</sup> It is estimated that two to three rounds would be needed to obtain consensus, and that each round will take 30 to 45 minutes to complete.<sup>12</sup> Planning for up to three rounds is thought to allow sufficient time for agreement to be reached about most items, as well as facilitate efficient and organised discussion of proposed items at the consensus meeting. Multiple rounds give participants the opportunity to revise their views and to identify which aspects of reporting they find truly important.

#### **4. Project Phase 3: Consensus Development Conference**

##### ***4.1 Objective***

The purpose of Phase 3 is to select the specific reporting items to be included in the CONSORT-SPI checklist. Though useful, the Delphi method has been criticised as potentially forcing consensus and by not allowing participants to fully discuss opinions raised.<sup>12</sup> A face-to-face meeting is proposed in order to help elucidate reasons for held opinions (which is harder to clarify without in-person interaction), and to consider the most important reporting items in greater depth.<sup>13</sup> Therefore, a consensus development process will be used for the actual selection of items for the CONSORT-SPI Extension so that participants can confirm the results of the Delphi process and discuss the opinions raised in greater detail. Based on previous reporting guideline projects, a 3-day formal consensus meeting will be held to determine guideline *content*, rather than wording or format.<sup>14</sup> This amount of time has been allotted to sufficiently allow for thorough discussion, reducing stressful and hasty decision-making that can hinder judgment.<sup>13</sup>

## ***4.2 Thesis Chapter 5: Protocol for Consensus Meeting Methods***

Chapter 5 of the thesis will consist of a report on the design, proceedings, and discussion during the consensus meeting. This report will serve as an initial foundation for the CONSORT-SPI checklist document. The DPhil candidate, along with other members of the Project Executive, contacted members of the IAG, set the agenda for the consensus meeting, and led discussions during the meeting. The DPhil candidate also created the first draft of the consensus meeting agenda, prepared all of the presentations at the meeting, and conducted the data analysis presented in this chapter.

### *4.2.1 International Advisory Group Meeting*

Members of the IAG will be individually consulted to decide: a list of potential consensus meeting participants, the duration of the meeting, logistics for the meeting and the meeting agenda, and materials to send prior to the meeting.

### *4.2.2 Selection of Participants*

Participants will be drawn from the Delphi process and from IAG suggestions, and selected purposively by the Project Executive to include a range of perspectives.<sup>42</sup> The size of the group—intended to be 20-30 participants—will aim to balance diversity of opinion with meaningful opportunities for interaction.<sup>2</sup> Compared to the Delphi process, the consensus meeting will be smaller in order to maximise the ability to achieve consensus.<sup>13</sup>

### *4.2.3 Structure*

The consensus meeting is proposed to follow established methods<sup>14</sup> used in previous CONSORT meetings.<sup>4,5,7,10</sup> Literature reviews and results of the Delphi process will be provided to participants in advance in order to ground conversations in empirical

information and thus facilitate cohesive discussion.<sup>13</sup> The meeting itself will involve presentations on relevant background literature, sharing the results of the Delphi process, discussion of each proposed checklist item, voting that follows from discussion, deciding what outputs need to be produced, and deliberating about the optimal dissemination and implementation strategy.

The conference will begin with several presentations about specific background topics related to reporting social and psychological intervention trials.<sup>7</sup> The results of the Delphi process will then be formally presented. Afterwards, participants will be led in structured discussions of each item proposed for the checklist from the Delphi process.<sup>7</sup> Informed by literature reviews, the Delphi process, and empirical evidence, participants will discuss whether each proposed checklist item should be included or modified for the final version of the checklist, expounded in the E&E only, or dropped from the guideline altogether.<sup>5</sup> The Project Executive will facilitate these discussions to ensure that all participants can express views, that all ideas are discussed in-depth, and that assertive participants do not dominate the discussion.<sup>13</sup>

Participants will discuss and vote on the minimum items to include in the guideline checklist. High ratings for inclusion will be required to ensure that the results represent general agreement of a substantial majority, and to produce a concise checklist.<sup>13</sup> Voting will be confidential using anonymous ballots to promote honest answers and allow participants to rethink their position if a re-vote is needed.<sup>21</sup> Minority and alternative views will be documented and may be noted in guideline documents.<sup>13</sup> The meeting will conclude with discussion about optimising dissemination, and participants will be asked to volunteer for specific efforts to this end.<sup>7</sup> The meeting will feed into guideline development by generating the list of items to be modified for or added to the CONSORT-SPI Extension checklist, as well as providing information for the E&E document.

## **5. Project Phase 4: Drafting the Guideline Documents**

After the consensus meeting, Phase 4 involves drafting the guideline documents so that their wording and content is clear, precise, and suitable for all disciplines.

### ***5.1 Thesis Chapter 6: Protocol for CONSORT-SPI E&E Template Development***

Chapter 6 of the thesis will constitute a full explication of the rationale for each reporting item in the CONSORT-SPI checklist. This chapter will serve as an initial template for the E&E document discussed below. This template has been created by the DPhil candidate under the guidance of his supervisors.

#### ***5.1.1 Draft Explanation and Elaboration Document***

Following the consensus meeting, the proposed CONSORT-SPI checklist will be revised by the Project Executive. It will then be circulated to the consensus meeting participants so that it accurately reflects meeting discussions; it will also be sent to members of the IAG not at the meeting for feedback. The first goal will be to draft a checklist using concise yet comprehensive wording, a process that will require several iterations.

For this checklist, an Explanation and Elaboration (E&E) Document will explain in-depth the scientific rationale for each recommendation and provide an example of clear reporting for each item. This additional document will help editors and authors understand the importance of these guidelines, students and researchers understand the relevant issues, and authors meet the guideline requirements.<sup>2</sup>

In this dissertation, a draft E&E template for the current CONSORT-SPI checklist will be constructed using information gathered during earlier phases of the project and bolstered by other key literature in social and psychological intervention research.

Examples of good reporting will be derived from trials in the Phase 1 systematic reviews, using quotations from RCTs in all of the disciplines sampled in order to provide examples across key disciplines. The rationale underpinning each item will be derived from previous guidelines identified in Phase 1, feedback from the Delphi process in Phase 2, and discussions during the consensus meeting in Phase 3.

## **6. Project Phase 5: Guideline Dissemination**

### ***6.1 Thesis Chapter 7: Protocol for the Development of the Dissemination Strategy***

Chapter 7 of the thesis will map several implementation frameworks within the discipline of “implementation science”<sup>16</sup> onto current and proposed guideline dissemination and implementation activities for CONSORT-SPI. This chapter will form the basis of the dissemination and implementation strategy for CONSORT-SPI. This draft has been created by the DPhil candidate under the guidance of his supervisors.

#### *6.1.1 Insights from Previous Reporting Guidelines and the CONSORT-SPI Consensus Meeting*

The goal of the dissemination and implementation plan is to maximise awareness, understanding, and use of the CONSORT-SPI Extension when reporting social and psychological intervention trials. To help achieve this end, this plan involves activities that span across all phases of the project. There are several widely-utilised dissemination techniques for reporting guidelines. As a result, there will be a multi-pronged dissemination strategy to ensure that all targeted stakeholders benefit from this guideline. In addition, an analysis of discussions at the consensus meeting about the activities for the dissemination strategy will help inform the draft of this plan in this chapter of the dissertation.

### *6.1.2 Review of Implementation Science for Clinical Guidelines*

To turn the proposed activities from previous guidelines and the consensus meeting into a coordinated dissemination and implementation plan, theoretical frameworks from the implementation science and clinical guideline literature will be analysed and applied to reporting guidelines. Implementation of reporting guidelines remains variable and often lacks a theoretical foundation, despite the numerous organisational and individual factors influencing the quality of research reporting that exist. It is plausible that multifaceted interventions built upon a careful assessment of barriers and a coherent theoretical base may be more effective than single implementation interventions.<sup>43</sup> While frameworks for implementation are being increasingly used for empirically supported interventions and clinical guidelines, no paper to date has applied them to reporting guidelines. Insights from the implementation of interventions and clinical guidelines have the potential to improve the uptake of reporting guidelines.

An investigation will be conducted by the DPhil candidate on implementation science literature related to the dissemination of clinical guidelines and empirically supported interventions, which will then be applied to reporting guidelines via CONSORT-SPI as a case study. This investigation will also suggest where future areas of empirical research are needed in the reporting guidelines field to assess best ways of implementing reporting guidelines. Several frameworks will be used, including the Theoretical Domains Framework (TDF),<sup>16</sup> the Guideline Implementability Framework (GIF),<sup>44</sup> the Promoting Action on Research in Health Services (PARiHS) framework,<sup>45</sup> and the Consolidated Framework for Implementation Research (CFIR).<sup>46</sup>

## **7. Thesis Chapter 8: Conclusion**

The thesis will conclude with a final chapter summarising the project, discussing its

implications, and offering suggestions for future research. The proposed methods are based on best practices and were chosen to develop the most appropriate reporting standards, generate consensus, and promote widespread dissemination and uptake of the CONSORT-SPI Extension. If executed successfully, the outputs from this project will help authors write clear reports, create a framework for reviewers to assess publications, expedite and strengthen funding evaluations, provide a pedagogical tool for understanding social and psychological intervention trials, and help consumers evaluate studies. In these ways, the guideline will improve the quality of primary research and facilitate the efficient, effective transfer of research evidence into real-world use.

## References

1. Montgomery P, Grant S, Hopewell S, et al. Protocol for CONSORT-SPI: An Extension for Social and Psychological Interventions. *Implementation Science*. 2013;8:99.
2. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
3. Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR Network's Survey of Guideline Authors. *PLoS Medicine*. 2008;5:869-874.
4. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698-702.
5. Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT statement by high impact factor medical journals: A survey of journal editors and journal 'Instructions to Authors'. *Trials*. 2008;9:20.
6. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Reporting randomized, controlled trials of herbal interventions: An elaborated CONSORT Statement. *Annals of Internal Medicine*. 2006;144:364-367.
7. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
8. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: Defining standard protocol items for clinical trials. *Annals of Internal Medicine*. 2013;158(3):200-207.

9. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: Extension to cluster randomised trials. *BMJ*. 2004;328:702-708.
10. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ*. 2008 337:a2390.
11. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.
12. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*. 2000;32(4):1008-1015.
13. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998;2(3).
14. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311(376).
15. Bowling A. *Research methods in health: Investigating health and health services*. New York: McGraw-Hill; 2009.
16. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Science*. 2012;7:37.
17. Moher D. Reporting research results: A moral obligation for all researchers. *Canadian Journal of Anesthesiology*. 2007;54:331-335.
18. Glasziou P, Chalmers I, Altman DG, et al. Taking healthcare interventions from trial to practice. *BMJ*. 2010;341:384-387.
19. Altman DG. Endorsement of the CONSORT statement by high impact medical journals: A survey of instructions for authors. *BMJ*. 2005;330:1056-1057.

20. Gorard S. What should we do in the absence of trials? *Trials in Public Policy*. 2006;2.
21. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*. 2011;8(1):e1000393.
22. McKenna HP. The Delphi technique: A worthwhile research approach for nursing? *Journal of Advanced Nursing*. 1994;19:1221-1225.
23. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technology Assessment*. 2004;8(6).
24. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *MJA*. 2006;185:263-267.
25. Stinson JN, McGrath PJ, Yamada JT. Clinical trials in the Journal of Pediatric Psychology: Applying the CONSORT statement. *Journal of Pediatric Psychology*. 2003;28 159-167.
26. Montgomery P, Grant SP. Development of a CONSORT extension for complex social interventions. Paper presented at: Meeting at the 20th annual Cochrane Colloquium. 2012, October; Auckland, New Zealand.
27. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *The Lancet*. 2008;371:281-283.
28. Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. A comparison of RCTs in health and education. *British Educational Research Journal*. 2005;31(6):761-785.
29. Gardner F, Mayo-Wilson E, Montgomery P, et al. Editorial Perspective: The need

- for new guidelines to improve the reporting of trials in child and adolescent mental health. *Journal of Child Psychology and Psychiatry*. 2013;54(7):810–812.
30. Grant S, Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo- Wilson E. Letter to the Editor: New guidelines are needed to improve the reporting of trials in addiction sciences. *Addiction*. 2013;108:1687-1688.
31. Montgomery P, Grant S, Hopewell S, Macdonald G, Moher D, Mayo-Wilson E. Developing a reporting guideline for social and psychological intervention trials. *British Journal of Social Work*. 2013;43(5):1024-1038.
32. Montgomery P, Mayo-Wilson E, Hopewell S, G M, Moher D, Grant S. Developing a reporting guideline for social and psychological intervention trials. *American Journal of Public Health*. 2013;103(10):1741-1746.
33. Grant S, Mayo-Wilson E, Hopewell S, Macdonald G, Moher D, Montgomery P. Developing a reporting guideline for social and psychological intervention trials. *Journal of Experimental Criminology*. 2013;9(3):355-367.
34. Grant S, Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo-Wilson E. Developing a reporting guideline for social and psychological intervention trials. *Research on Social Work Practice*. 2013;23(6):595-602.
35. Mayo-Wilson E, Grant S, Hopewell S, Macdonald G, Moher D, Montgomery P. Developing a reporting guideline for social and psychological intervention trials. *Trials*. 2013;14:242.
36. Mayo-Wilson E, Montgomery P, Hopewell S, Macdonald G, Moher D, Grant S. Developing a reporting guideline for social and psychological intervention trials. *British Journal of Psychiatry*. 2013;203:250-254.
37. Gawande A. *The checklist manifesto: How to get things right*. London: Profile Books; 2010.

38. Dwan K, Gamble C, Williamson PR, Altman DG. Reporting of clinical trials: A review of research funders' guidelines. *Trials*. 2008;9(66).
39. Marsh P, Fisher M. *Developing the evidence base for social work and social care practice*. London: Social Care Institute for Excellence; 2005.
40. Littell JH, Corcoran JC, Pillai V. *Systematic reviews and meta-analysis*. New York, NY: Oxford University Press; 2008.
41. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: Explanation and Elaboration. *PLoS Medicine*. 2008;5(1):e20.
42. Moher D, Weeks L, Ocampo M, et al. Describing reporting guidelines for health research: A systematic review. *Journal of Clinical Epidemiology*. 2011;64(7):718-742.
43. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. *Implementation Science*. 2012;7:50.
44. Gagliardi AR, Brouwers M, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. *Implementation Science*. 2011;6:26.
45. Stetler CB, Damschroder LJ, Helfrich CD, Hagedorn HJ. A guide for applying a revised version of the PARIHS framework for implementation. *Implementation Science*. 2011;6:99.
46. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*. 2009;4:50.

### Chapter 3:

#### **Reporting Quality of Social and Psychological Intervention Trials: A Systematic Review of Reporting Guidelines and Trial Publications**

*NB: This chapter has been published in an alternative form in PLoS One<sup>1</sup>*

#### **Abstract**

**Objective:** To assess reporting guidelines for and the reporting quality of social and psychological intervention trial publications.

**Design:** A two-part study was conducted: (1) a systematic review of reporting guidelines and reporting quality assessment tools related to social and psychological intervention trials, and (2) a review of the reporting quality of social and psychological intervention trials.

**Data Sources:** To identify reporting guidance, a peer-reviewed search strategy was used across multiple electronic databases. Reporting guideline registries were also examined. To identify trial publications, a hand-search was conducted on the 10 journals with the highest impact factors each in clinical psychology, criminology, education, public health, and social work.

**Eligibility Criteria:** Reporting guidance had to consist of a peer-reviewed article that introduced a reporting guideline or quality assessment tool with standards relevant to experimental evaluations of social and psychological interventions. Social and psychological intervention trials had to report a randomised experiment that evaluated psychological, social, and/or health outcomes.

**Results:** The systematic review of reporting guidance identified 14 reporting guidelines and 5 reporting quality assessment tools, yielding a total of 147 reporting standards.

Developers of official CONSORT guidance used more recommended techniques in reporting guideline development and dissemination than developers of other guidelines. Guidelines developed by behavioural and social scientists yielded 89 modified standards for social and psychological interventions not found in CONSORT guidance. The review of social and psychological intervention trials (n = 309) revealed that many aspects of RCTs currently addressed in CONSORT guidelines were poorly reported. Authors rarely provided information related to blinding (14.7%), random sequence generation (24.4%), and allocation concealment (16.3%). Other important items were also poorly reported: few studies included aspects of the actual delivery by providers of the intervention (40.7%) and control (30.9%) treatments, uptake by participants of the intervention (24.0%) and control (19.0%) treatments, and the timing and setting of the trials (47.0%). Only 14 of 50 journals referenced a reporting guideline related to trials in their “Instructions to Authors.”

**Conclusion:** Current guidelines relevant to social and psychological intervention trials demonstrate limitations in their content, development, and dissemination. Moreover, important details about social and psychological intervention trials are poorly reported. Our results underscore the need for a new CONSORT extension specifically for social and psychological interventions.

## 1. Introduction

The theoretical and empirical literature discussed in Chapter 1 suggests the utility of a new reporting guideline for social and psychological intervention trials. Namely, across key disciplines—such as criminology, education, psychology, public health, and social work—social and psychological interventions are increasingly evaluated via RCTs in order to inform policy and practice decision-making. These interventions are often complex and delivered in environments that are difficult to control and to measure, which complicates assessments of trial quality as well as interpretations about the generalisability of trial results.<sup>2,3</sup> High quality reports of social and psychological intervention trials are important to diverse groups of stakeholders, including researchers, journal editors, funding agencies, practitioners, policy-makers, and research participants. These research consumers depend on accurate, complete, and transparent reports to appraise the validity and generalisability of trials.

To address these needs, researchers and journal editors have developed reporting guidelines<sup>4</sup> that highlight key information about internal validity, external validity, and knowledge transfer of trials (e.g., locating trials in databases, assessing conflicts of interest). As mentioned in Chapter 1, the CONSORT Statement and its extensions are the preeminent guidelines for reporting trials.<sup>5</sup> Despite the impact of CONSORT guidelines,<sup>6</sup> several studies suggest that deficiencies persist in the reporting of social and psychological intervention trials.<sup>7-11</sup> A common explanation is that current standards in prominent reporting guidelines are not adequately tailored to these trials, leading to less uptake and implementation of reporting standards in social and behavioural sciences compared with biomedical disciplines. A new guideline may be needed that requests more information related to social and psychological interventions, such as intervention theory of change, assessment of intervention mechanisms during the trial, and relevant information about the

influence of trial context.<sup>12-14</sup>

Before developing a new reporting guideline or extending an existing one, experienced reporting guideline developers, including members of the CONSORT Group and the Enhancing the QUALity and Transparency Of health Research (EQUATOR) Network, recommend a few preliminary actions to justify such a project.<sup>4,15</sup> Namely, a structured approach to reporting guideline development should begin with a needs assessment that (1) reviews whether an adequate guideline already exists for a given research method and (2) obtains evidence on the reporting quality of published research of that method.<sup>15</sup> A systematic review of the development, content, and dissemination of previous reporting guidelines would indicate whether any single existing guideline is already sufficient, or whether a new reporting guideline is truly needed for a given research area. Chapter 2 has proposed a project plan that follows these recommended techniques in guideline development and dissemination.<sup>16</sup>

Prior to the reviews reported in this chapter, no study had identified and assessed all reporting guidelines related to social and psychological intervention RCTs.<sup>16</sup> Moreover, though informative, previous reviews of the reporting quality of social and psychological intervention trials have two primary limitations: each individual review (a) focused on RCTs within one particular discipline and (b) assessed reporting quality according to only a sub-set of known reporting standards.<sup>7-11</sup> A larger review of social and psychological intervention trials across key disciplines and according to a comprehensive set of reporting standards is needed to clearly demonstrate deficiencies in reporting. Such a review would have the ability to suggest that the overall reporting of social and psychological intervention RCTs is inadequate, and that a single, unified set of reporting standards for these trials is feasible and needed.

This chapter reports a two-part study that assessed whether there is a need for a

new reporting guideline for social and psychological intervention trials. This study examined:

1. the quality of current reporting guidance relevant to social and psychological intervention trials; and
2. the current reporting quality of social and psychological intervention trials across several disciplines according to a comprehensive set of reporting standards for these studies.

## **2. Methods**

### ***2.1 Eligibility Criteria***

To be included in the first review on reporting guidelines, peer-reviewed, published reporting guidance must: (a) introduce a reporting guideline or reporting quality assessment tool with (b) standards for reporting experimental evaluations of social and psychological interventions. A reporting guideline or assessment measure was defined as a checklist, flow diagram, and/or explicit text containing a finalised, codified list of standards for reporting a particular research method.<sup>17</sup> Articles were excluded if they (a) introduced a guideline for the quality of RCT *design and conduct* rather than standards for *reporting* RCTs, (b) included reporting standards that were for a specific type of research or intervention design irrelevant to social and psychological intervention research (e.g., acupuncture trials), (c) were secondary publications of a reporting guideline or quality assessment tool (e.g., a review on RCT reporting quality that utilised an unmodified version of the CONSORT Statement), or (d) were commentaries or editorials about reporting standards. No other eligibility criteria were utilised.

To be included in the second review on RCT reporting quality, studies must have reported a randomised experiment of an intervention targeting psychological, social, and/or

health outcomes. Studies were excluded if they: (1) only reported RCT process evaluations without trial outcomes, (2) evaluated cost-effectiveness of an intervention, (3) used randomisation to balance order of exposure to conditions that were experienced by all participants, or (4) explicitly were evaluating any biomedical treatments (i.e., those specifically targeting biological mechanisms) or pharmacological techniques (i.e., interventions that employ drugs) in at least one trial group. Multiple studies reported in the same research article were considered as separate studies. No other eligibility criteria were used.

## ***2.2 Search Strategy and Study Selection***

In August 2011, the DPhil candidate used an adapted version of a peer-reviewed electronic search strategy<sup>17</sup> to identify reporting guidelines relevant to social and psychological intervention RCTs (see Box 1). Three registries of reporting guidelines and reporting quality assessment tools were also searched. Firstly, the EQUATOR Network is an international initiative that seeks to improve the reliability and utility of health research literature by promoting transparent and accurate reporting of research studies.<sup>4</sup> This initiative has developed and maintains a library of identified health research reporting guidelines. Secondly, a review on the development and contents of reporting guidelines for health research offers a catalogue of reporting guidelines.<sup>17</sup> Thirdly, a systematic review on studies that assessed the quality of conducting or reporting RCTs offers a register of reporting quality assessment tools.<sup>18</sup> Any eligible guidelines identified through this process also had their reference sections searched for further eligible studies. The current review did not include a grey literature search, as the purpose was to examine reporting standards that have gone through peer-review and have been codified into a published, available checklist.

**Box 1. Electronic Search Strategy Adapted from Moher 2011**

## I. Ovid MEDLINE(R) –1948 to June Week 1 2011

1. exp Research Design/
2. exp Guideline/
3. exp Study characteristics/
4. exp Epidemiologic studies/
5. Feasibility Studies/
6. Intervention Studies/
7. Program Evaluation/
8. Evidence-Based Medicine/
9. Human Experimentation/
10. exp Research/
11. ((control\$ or clinical or comparative\$) adj2 (trial\$ or stud\$)).mp.
12. between group design\$.mp.
13. random\$.tw.
14. ((control\$ or intervention or evaluation or comparative or effectiveness or evaluation or feasibility) adj3 (trial or studies or study or program or design)).tw.
15. systematic review\$.tw.
16. (time adj series).tw.
17. (pre test or pretest or posttest or post test).tw.
18. controlled before.tw.
19. or/1—18
20. (reporting and (trial\$ or studies)).ti.
21. ((guideline\$ or guide line\$ or checklist\$ or recommendation\$ or standard\$ or requirement\$ or instruction\$ or guidance\$ or policies or policy) adj3 (reporting or

publishing or good practice or good practice).tw.

22. or/20—21
23. (good adj3 practice\$ adj3 (reporting or publishing or publication)).tw.
24. (reporting adj2 (guideline or standard or standards or quality)).ti.
25. (19 and 22) or 23 or 24
26. limit 25 to english

## II. Ovid EMBASE—1980-2011, Week 22

1. exp Research/
2. Epidemiology/
3. Randomized Controlled Trial/
4. exp Clinical Trial/
5. Meta Analysis/
6. Systematic Review/
7. Evidence Based Practice/
8. Practice Guideline/
9. or/1—8
10. (reporting and (trial\$ or study or studies)).ti.
11. ((guideline\$ or guideline\$ or checklist\$ or checklist\$ or recommendation\$ or standard\$ or requirement\$ or instruction\$ or guidance\$ or policies or policy) adj3 (reporting or publishing or good practice or good practice).tw.
12. (or/10—11) and 9
13. (reporting adj2 (guideline or standard or standards)).ti.
14. (good adj3 practice\$ adj3 (reporting or publishing or publication)).tw.
15. or/12—14

16. limit 15 to english

C. Ovid PsycINFO—1806 to June, Week 1, 2011

1. Treatment Effectiveness Evaluation/
2. exp Clinical Trial/
3. Meta Analysis/
4. exp Experimentation/
5. Evidence Based Practice/
6. exp Experimental Design/
7. Methodology/
8. or/1—7
9. (reporting and (trial\$ or studies or study)).ti.
10. ((guideline\$ or guide line\$ or checklist\$ or check list\$ or recommendation\$ or standard\$ or requirement\$ or instruction\$ or guidance\$ or policies or policy) adj3 (reporting or publishing or good practice or good practi#e\$)).tw.
11. or/9—10
12. (reporting adj2 (guideline or standard or standards)).ti.
13. (good adj3 practi#e\$ adj3 (reporting or publishing or publication)).tw.
14. (11 and 8) or 12 or 13
15. limit 14 to english

## D. Cochrane Methodology Register—Cochrane Library 2011

Record title:

(reporting near/2 (guideline or standard or standards))

OR

Search all text:

((guideline\* or checklist\* or recommendation\* or standard\* or requirement\* or instruction\* or guidance\* or policies or policy) near/3 (reporting or publishing or good practi#e\*))

OR

Keywords:

(checklists and guidelines)

## E. Scopus Social Sciences—1966 to June, Week 1 2011

TITLE-ABS-KEY((reporting W/2 guideline) OR (reporting W/2 standard\*) OR (reporting near/2 quality))

## F. ISI Web of Knowledge Social Science Citation Index—1945 to June 7 2011

TS=((reporting near/2 guideline) OR (reporting near/2 standard\*) OR (reporting near/2 quality))

To locate RCTs of social and psychological interventions, the DPhil candidate conducted a hand search of the Table of Contents of journals publishing social and psychological intervention RCTs. These journals were identified from the ISI Web of Knowledge 2010 Journal Citation Reports (JCR) for Social Sciences, which at the time was the most recent JCR providing impact factors for social science journals. To obtain an

extensive sample of trials, the 10 highest impact factor journals in clinical psychology, criminology, education, public health, and social work (50 journals total) that published RCTs in 2010 were retrospectively searched for and analysed. To reflect impact factor rankings from the 2010 JCR, all social and psychological intervention RCTs published from January to December 2010 in these journals were eligible for inclusion.

### ***2.3 Data Extraction***

Reporting guidance was assessed by the DPhil candidate according to recommended techniques for developing and disseminating reporting guidelines,<sup>16,17</sup> which fell into four general domains: preliminary stages of development, guideline development, publication, and dissemination (see Appendix C). For each guideline, citations were counted through November 2012 using Google Scholar, which provides a wide measure of impact across social and behavioural science subfields and publication mediums.<sup>19</sup> If a guideline was published in multiple journals and/or had an official explanatory document detailing how to adhere to its reporting standards, citations were combined for all of these documents.

Reporting standards from all identified guidance were then compiled into a comprehensive, non-redundant, itemised list of standards relevant to social and psychological intervention RCTs. To note, the items comprising a reporting guideline or assessment tool often contain multiple reporting standards, each of which can be evaluated independently. For example, CONSORT item 7 requires that authors report (a) how sample size was determined and (b) an explanation of any interim analyses and stopping rules.<sup>5</sup> While this format is helpful to authors and journal editors in checklists, it is problematic for constructing a comprehensive list of individual reporting standards for the purposes of this review.<sup>20</sup> Therefore, reporting items in guidelines or assessment tools were

divided into any distinct sub-items so that single standards could be independently identified. Similar to previous studies, guideline content was analysed by mapping identified reporting standards onto items included in the CONSORT Statement in order to organise the checklist according to the common sections of a trial report (i.e. introduction, methods, results, discussion, and other information).<sup>7,20</sup> Any identified reporting standards in social and behavioural science guidelines that are not in official CONSORT guidelines were also noted.

For the review of reporting quality, the DPhil candidate first collected information about the selected journals, including their impact factors, discipline as listed in the JCR, and number of eligible RCTs in 2010. Akin to previous studies,<sup>15</sup> a full-text review was done on each journal's public "Instructions to Authors" webpages to identify whether the journals provided any guidance for reporting RCTs (e.g., instructions on the webpage, reference to a reporting guideline) and whether journals required trial registration prior to publishing the trial manuscript.

After determining all eligible social and psychological intervention RCTs, the DPhil candidate then assessed the reporting quality of eligible RCTs using the comprehensive checklist of reporting standards from the review of reporting guidelines (see Appendix D). For quality assurance, a second coder—a then-MPhil student serving as a research assistant—independently assessed the reporting quality of RCTs identified in clinical psychology, criminology, education, and social work journals (77% of the sample). The RCTs in public health journals were not double-coded, as the decision to include these studies in this review was made in light of feedback from the DPhil candidate's transfer of status examination, which was held after the double-coding and publication of that sub-set of the data from the other four disciplines.<sup>1</sup>

Each reporting standard was assigned a "yes" (score of 1) or "no" (score of 0)

response depending on whether the authors reported the information requested. The reviewers conducted preliminary analyses on four published studies—one each in clinical psychology, criminology, education, and social work—to gain familiarity with the tool and increase consistency of interpretation. Following these preliminary analyses, coding rules were adapted from previous studies about reporting quality.<sup>21</sup> Firstly, each sub-item aimed to address a single reporting standard that can be scored by a reviewer who is not an expert in the field. Secondly, one sentence in a trial report that contains multiple pieces of information can achieve compliance to several reporting standards simultaneously. In contrast to previous studies, if no primary outcomes were specified, all outcomes in a trial report were considered secondary outcomes, and all items referring to primary outcomes were coded as 0. It was agreed that not penalising would under-emphasise the importance of specifying primary outcomes for testing and delineating theory, in addition to their role in power calculations.<sup>22</sup> Lastly, unless specifically required by a reporting item, credit was granted if the authors provided the information in a different section of the paper than is listed in the CONSORT checklist. For example, Items 1a-1p must be reported in the title or abstract, but information on allocation could be in the methods, results, or discussion section.

After data had been extracted, discrepancies on the double-coded studies were resolved by discussion and consensus. Using SPSS version 18, inter-rater agreement prior to discussion was calculated as  $\kappa = 0.71$ , indicating substantial agreement.<sup>23</sup> Data resolved after discussion were used for the final analyses.

#### ***2.4 Data Analysis***

Data about adherence of reporting guidance to recommended techniques and of RCTs to reporting standards were summarised as frequencies,<sup>17</sup> and total citations of each

guideline were converted into median citations per year. Data about the development and dissemination of guidelines were compared by the following pre-specified types of reporting guideline: official CONSORT guidelines, non-CONSORT guidelines for medical sciences, or non-CONSORT guidelines for social and behavioural sciences.

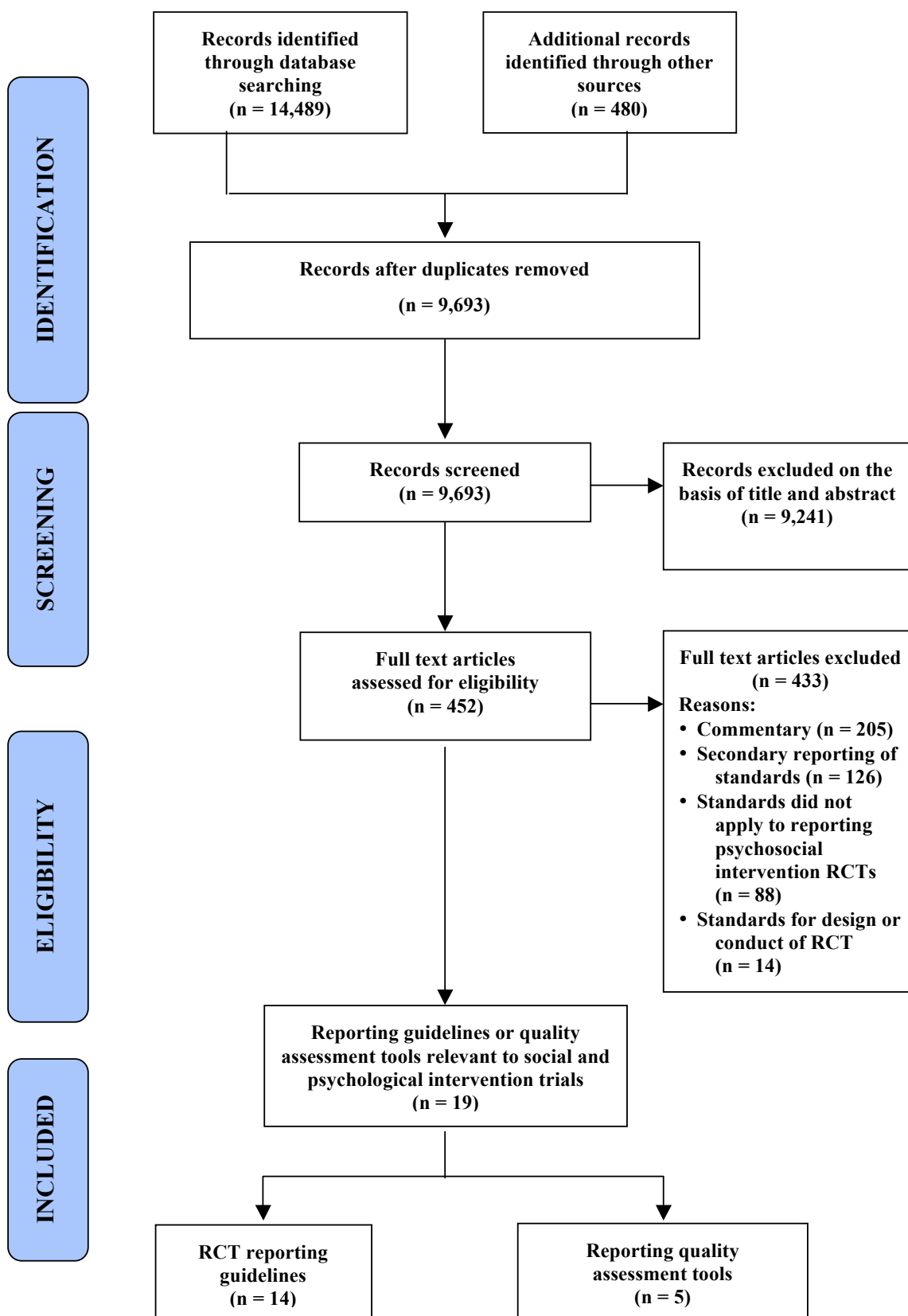
The reporting standards identified through this review are listed qualitatively in table format (see Appendix D). Compliance to reporting standards was analysed for the whole sample and by academic discipline to provide a preliminary view of differences in reporting across social and behavioural sciences. Reporting items were then categorised into *a priori* conceptual groupings: internal validity, external validity, and other important study details. “Study details” include information in the title and abstract for indexing in electronic databases, reporting certain ethical concerns (e.g., ethics board approval), and if authors provided manuals or protocols for the study as well as the intervention. Items related to the Cochrane Risk of Bias tool were also specifically identified. This tool is used when conducting a systematic review to assess the chance that methods used in an RCT may lead to invalid or biased results.<sup>24</sup> Clear reporting is essential to making such assessments.

### **3. Results**

#### ***3.1 Previous Guidance for Social and Psychological Intervention Trials***

The flow of identifying eligible reporting guidance can be found in Figure 1. After removing all duplicates, 9,693 studies were identified. After 9,241 records were excluded by screening articles’ titles and abstracts, a full-text review of 452 studies yielded 19 eligible articles. The search identified 14 eligible reporting guidelines and 5 eligible reporting quality assessment tools (see Table 1). All documents were developed between 1980 and 2010 (median = 2004). Six reporting guidelines were officially developed by the

Figure 1. Flowchart of Records through Literature Search for Reporting Guidance



CONSORT Group. Three assessment tools and three reporting guidelines were non-CONSORT documents for health research in general. Three assessment tools and four reporting guidelines were specific to the social and behavioural sciences, such as behavioural change interventions,<sup>25,26</sup> education,<sup>27</sup> psychology,<sup>28</sup> and criminal justice.<sup>7</sup>

Overall, the CONSORT guidelines used recommended techniques for all stages of guideline development and dissemination at higher percentages than non-CONSORT documents in medical, social, and behavioural sciences (see Table 2). Notably, CONSORT guidelines reported using more recommended techniques (e.g., rigorous consensus methods) in the development stage (75.0%) than other medical guidelines (44.4%) and social and behavioural science guidelines (31.0%). CONSORT guidelines also adhered to more dissemination activities (76.7%), such as endorsement and adherence by journals, than other medical guidelines (10.0%) and social and behavioural science guidelines (37.1%). In addition, CONSORT guidelines had considerably higher citations per year (73.7) than other guidelines in medicine (9.9) and the behavioural and social sciences (4.4). A breakdown of the specific activities performed by each guideline team can be found in Appendix E.

The 19 included reporting guidelines included a median of 32 reporting standards (interquartile range (IQR) = 17 to 54; range=3 to 201). From these, a checklist was developed of 147 non-redundant reporting standards that are relevant to social and psychological interventions. Similar to previous studies,<sup>7,20</sup> these standards were organised according to the 25 items of the CONSORT Statement (see Appendix D). While no current guidance was specifically developed for or sufficiently incorporated all key standards for social and psychological intervention trials, 89 distinct reporting standards were identified that are either not currently found in CONSORT guidelines or are tailored versions of CONSORT items for social and behavioural sciences (see Appendix F). These standards

**Table 1. Characteristics of Included Reporting Guidelines and Quality Assessment Tools**

<b>Guideline</b>	<b>Year</b>	<b>Document Type</b>	<b>Official CONSORT</b>	<b>Targeted Area</b>	<b># Reporting Standards</b>
Reporting Guidelines Specific to the Social and Behavioural Sciences					
Alcohol Outcome Studies Coding Sheet [26]	2010	AT		Alcohol	36
AERA Standards for Empirical Social Science Research [27]	2006	RG		Education	56
CONSORT and Criminal Justice Trials (CJT) Project Coding Sheet [9]	2010	AT		Criminology	54
Journal Article Reporting Standards [41]	2008	RG		Psychology	134
Nelson-Moberg Expanded CONSORT Instrument [34]	2004	AT		Occupational Therapy	201
TREND Statement [38]	2004	RG		Public Health	59
WIDER [39]	2009	RG		Behavioural Change Interventions	12
Other Reporting Guidelines					
CONSORT Extension for Abstracts [28]	2008	RG	x	Abstracts	17
CONSORT Extension for Cluster Trials [29]	2004	RG	x	Cluster Trials	40
CONSORT Extension for Non-Pharmacological Treatments [30]	2008	RG	x	Non-Pharmacological Interventions	27
CONSORT Extension for Pragmatic Trials [31]	2008	RG	x	Pragmatic Trials	25
CONSORT Extension for Reporting Harms [32]	2004	RG	x	Harms	22
CONSORT Statement [6]	1996	RG	x	None	37
Evidence-Based Behavioral Medicine-Specific Guidelines [33]	2003	RG		Behavioural Medicine	34
Jadad Scale [40]	1996	AT		None	3
Oxford Implementation Index [35]	2007	AT		Complex Interventions	17
Quality Evaluation Form [36]	1995	AT		None	20
Reporting Standards for Controlled Trials [42]	1980	RG		None	6
Structured Reporting of Randomized Controlled Trials [37]	1994	RG		None	32

In "Document Type" column, AT = reporting quality assessment tool, and RG = reporting guideline. In "Official CONSORT" column, a "x" means that the guideline is an official CONSORT guideline.

doi:10.1371/journal.pone.0065442.t001

**Table 2. Average Percentage of Employed Recommended Practices for Guideline Development by Document Type**

<b>Guideline Development Stage</b>	<b>CONSORT (n = 6)</b>	<b>Non-CONSORT Medical (n = 6)</b>	<b>Social &amp; Behavioural Science (n = 7)</b>
1. Preliminary Activities	91.7%	70.8%	67.9%
2. Document Development	75.0%	44.4%	31.0%
3. Publication Strategy	66.7%	5.5%	23.8%
4. Dissemination	76.7%	10.0%	37.1%
Median Citations per Year ( <i>Range</i> )	73.7 (43.3 – 535.5)	9.9 (0.2 – 480.2)	4.4 (1.0 – 65.0)

Citation count derived from Google Scholar search on 1 November 2012.  
 Stage 1 = 4 items, Stage 2 = 6 items, Stage 3 = 3 items, Stage 4 = 5 items  
 doi:10.1371/journal.pone.0065442.t002

mapped onto 18 CONSORT items: intervention theory of change (item 2), the implementation and actual uptake of interventions (item 5), properties of measurement (item 6), appropriate analysis of process and outcome data (items 12 and 17), generalisability of the trial population and setting to other contexts (items 4, 21, and 22), and conflicts of interest and ethical concerns for social and psychological interventions (item 25).

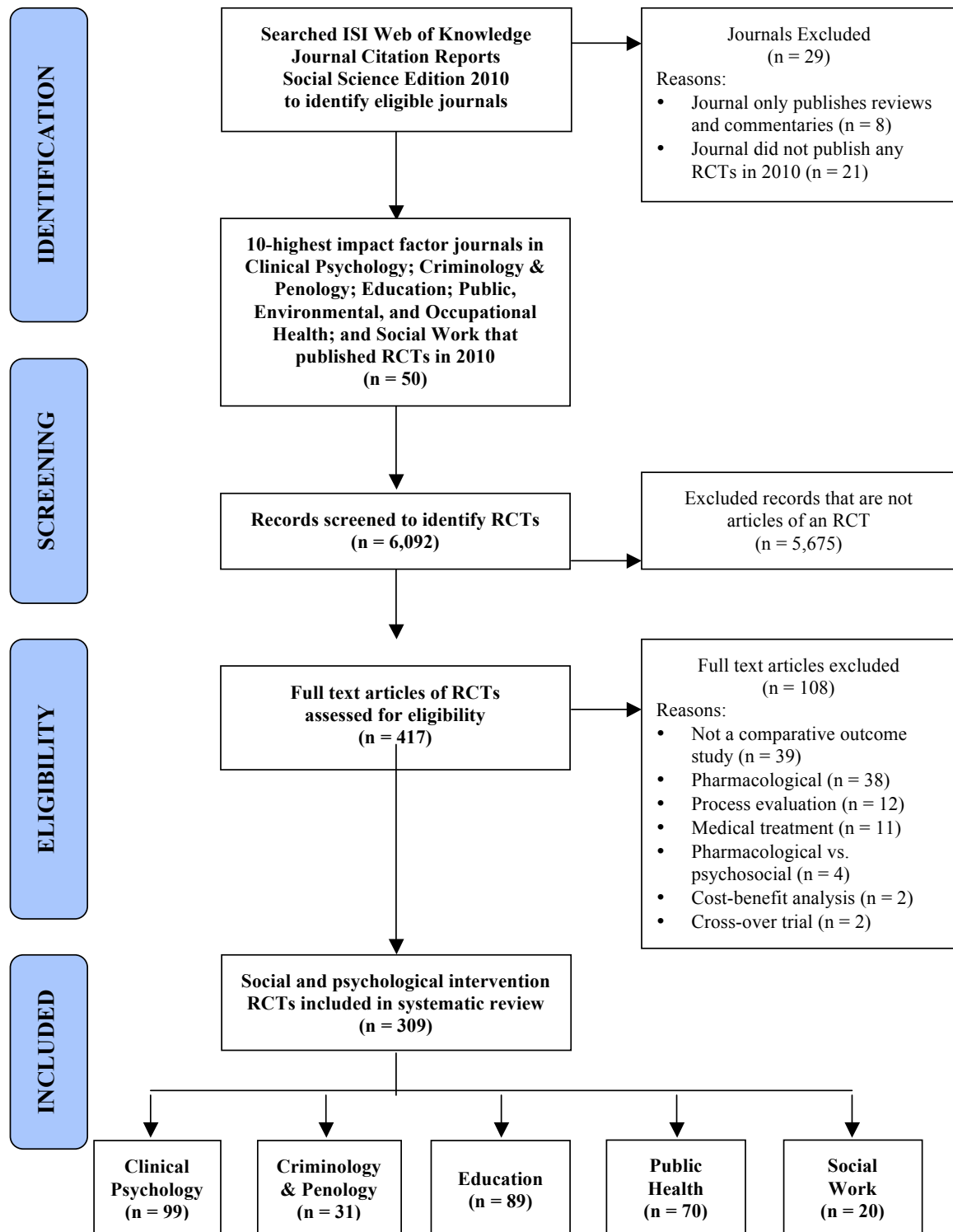
### ***3.2 Assessment of Reporting Quality of Social and Psychological Intervention Trials***

For the review of trial reporting quality, the literature search yielded 309 eligible trials (Figure 2); please see Appendix G for a list of all the trials. The number of eligible trials in 2010 ranged from 1 to 39 per journal, with a median of 3 trials per journal. Journal impact factor ranged from 0.792 to 5.235, with a median of 2.349 (see Table 3). Only 14 of the 50 journals (28%) referenced a reporting guideline in their “Instructions to Authors” section, while an additional 2 journals (4%) did provide advisory text about reporting certain aspects of intervention studies—without reference to any published reporting guideline. No other journals provided any textual guidance specific to reporting trials. Only 9 journals (18%) required trials to be registered in a trial registry (e.g., [clinicaltrials.gov](http://clinicaltrials.gov)) prior to publication.

#### ***3.2.1 Reporting Characteristics: General Reporting Quality.***

Overall, trials reported a mean of 41.9% of all reporting standards in the full sample; compliance with reporting standards related to internal validity (38.3%), external validity (45.3%), and study details (36.2%) was low (Table 4). On average, trials reported 48.3% of standards for implementation data about the intervention group, and 36.7% of standards for data about the control group. Trials inconsistently complied with several

Figure 2. Flowchart of Considered Publications through Review of Reporting Quality



items needed to use the Cochrane Risk of Bias tool; on average, trials reported 24.4% of standards related to random sequence generation, 16.3% of standards related to allocation concealment, 14.7% of standards related to blinding, and less than half of standards related to incomplete outcome reporting, such as standards for data analysis (42.8%) and participant flow (38.2%). Reporting quality did not vary much by discipline. Adherence to reporting items related to internal validity ranged on average from 27.4% (trials in education) to 47.0% (trials in psychology), external validity from 40.2% (trials in public health) to 48.4% (trials in psychology), and study details from 20.3% (trials in education) to 44.1% (trials in public health). Please see Appendix H for the reporting quality scores for each item by discipline.

### *3.2.2 Reporting Characteristics: Title, Abstract, and Introduction*

Only 25.2% of articles identified the study as a randomised trial in the title, though almost two-thirds (63.8%) identified the study as randomised in the abstract. The majority of studies did adhere to published standards in the introduction, with almost all of the articles (98.1%) reporting the objective of the study.

### *3.2.3 Reporting Characteristics: Methods*

Reporting quality of the methods was variable. Trials generally identified the type of trial (e.g., cluster or individual randomisation; 99.0%), named the experimental (100%) and control (94.5%) interventions, and provided the follow-up period of outcome evaluations (96.4%). Trials did not consistently include information about the implementation of intervention and control treatments: studies reported according to a mean of 40.7% of standards related to actual delivery by providers in the intervention group and 30.9% of standards in control groups, and a mean of 24.0% of standards related

to uptake by intervention group participants and 19.0% of standards related to uptake by control participants. Trials reported a mean of 29.8% of standards related to the wider service environment and 26.9% of standards related to the specific delivering organisations. Overall, 64.1% of trials reported the eligibility criteria, although the majority of these studies did not explicitly list all inclusion *and exclusion* criteria, the reasons why these criteria were selected, and implications for generalisability.

#### *3.2.4 Reporting Characteristics: Results*

While the majority of papers (75.4%) reported the number of participants randomised to each condition, few papers reported other aspects of participant flow through the trial, such as the number of participants: approached (38.2%), eligible for the study (38.2%), completing treatment (32.0%), and in primary analyses (35.9%). Less than half of the studies reported primary outcomes (28.8%) or secondary outcomes (49.8%) sufficiently to be included in meta-analyses. Very few studies (6.1%) indicated trial registration as well. Less than half of the studies reported other items important for social and psychological interventions, including incentives for participants (48.2% of studies), socio-demographic baseline characteristics (49.5% of studies), and adverse events (10.4% of studies).

#### *3.2.5 Reporting Characteristics: Discussion*

Trials reported on average 82.8% of standards regarding discussion of the limitations and 68.3% of standards related to generalisability of the study, as well as a mean of 83.2% of standards related to how trial findings fit with other evidence from previous studies. Fewer trials made reference to systematic reviews related to the intervention (53.7% of studies) and other published reports of the trial (36.9% of studies).

**Table 3. Sample of Journals Included in Reporting Quality Review**

<b>Journal</b>	<b>ISI 2010 Impact Factor</b>	<b>Reporting Guidance Specific to RCTs in Instructions to Authors</b>	<b>Trial Registration Required</b>	<b>Eligible RCTs in 2010</b>
<b>Clinical Psychology</b>				
<i>Archives of Sexual Behavior</i>	3.660	No	No	2
<i>Health Psychology</i>	3.982	CONSORT; JARS	Yes	16
<i>Journal of Abnormal Child Psychology</i>	3.564	No	No	7
<i>Journal of Abnormal Psychology</i>	5.235	JARS	No	1
<i>Journal of Behavioral Medicine</i>	3.232	No	No	14
<i>Journal of Clinical Child and Adolescent Psychology</i>	3.440	CONSORT; JARS	Yes	8
<i>Journal of Clinical Psychiatry</i>	5.023	Text about reporting intervention studies	Yes	5
<i>Journal of Consulting and Clinical Psychology</i>	5.227	JARS	No	35
<i>Neuropsychology</i>	3.176	CONSORT; JARS	Yes	2
<i>Psychological Medicine</i>	5.200	No	No	9
<b>Criminology</b>				
<i>British Journal of Criminology</i>	1.612	No	No	1
<i>Crime &amp; Delinquency</i>	1.750	No	No	1
<i>Criminal Justice and Behavior</i>	1.590	No	No	4
<i>Criminology</i>	2.658	No	No	1
<i>International Journal of Offender Therapy and Comparative Criminology</i>	1.071	No	No	2
<i>Journal of Criminal Justice</i>	1.076	No	No	3
<i>Journal of Interpersonal Violence</i>	1.354	No	No	6
<i>Justice Quarterly</i>	1.211	No	No	1

<b>Journal</b>	<b>ISI 2010 Impact Factor</b>	<b>Reporting Guidance Specific to RCTs in Instructions to Authors</b>	<b>Trial Registration Required</b>	<b>Eligible RCTs in 2010</b>
<b>Criminology</b>				
<i>Psychology, Crime &amp; Law</i>	1.133	No	No	11
<i>Youth Violence and Juvenile Justice</i>	1.132	No	No	1
<b>Education</b>				
<i>American Educational Research Journal</i>	2.479	AERA	No	3
<i>Computers &amp; Education</i>	2.617	No	No	39
<i>Early Childhood Research Quarterly</i>	2.192	Text about reporting effect sizes	No	4
<i>Educational Evaluation and Policy Analysis</i>	1.919	AERA	No	1
<i>Journal of Engineering Education</i>	2.219	No	No	7
<i>Journal of Research in Science Teaching</i>	2.728	No	No	7
<i>Journal of Teacher Education</i>	1.891	No	No	3
<i>Learning and Instruction</i>	2.768	No	No	19
<i>Metacognition and Learning</i>	2.038	No	No	2
<i>Science Education</i>	1.900	No	No	4
<b>Public Health</b>				
<i>American Journal of Public Health</i>	3.850	CONSORT; TREND	Yes	18
<i>Scandinavian Journal of Work Environment &amp; Health</i>	3.540	CONSORT, TREND	Yes	1
<i>AIDS and Behavior</i>	3.195	No	No	16
<i>Journal of Adolescent Health</i>	3.116	No	No	12
<i>Tobacco Control</i>	3.077	CONSORT, EQUATOR	Yes	1
<i>Public Health Genomics</i>	3.049	No	No	1

<b>Journal</b>	<b>ISI 2010 Impact Factor</b>	<b>Reporting Guidance Specific to RCTs in Instructions to Authors</b>	<b>Trial Registration Required</b>	<b>Eligible RCTs in 2010</b>
<b>Public Health</b>				
<i>Journal of Epidemiology and Community Health</i>	2.983	EQUATOR Network	No	3
<i>Nicotine &amp; Tobacco Research</i>	2.801	No	Yes	5
<i>Prevention Science</i>	2.754	No	No	8
<i>Social Science &amp; Medicine</i>	2.742	No	No	5
<b>Social Work</b>				
<i>American Journal of Community Psychology</i>	1.722	JARS	No	1
<i>Child Abuse &amp; Neglect</i>	1.945	No	No	2
<i>Child Maltreatment</i>	1.984	No	No	2
<i>Children and Youth Services Review</i>	1.130	No	No	3
<i>Family Relations</i>	1.216	No	No	2
<i>Health &amp; Social Care in the Community</i>	1.008	CONSORT; TREND	Yes	1
<i>Health &amp; Social Work</i>	1.143	No	No	1
<i>Journal of Community Psychology</i>	0.792	No	No	1
<i>Research on Social Work Practice</i>	1.130	JARS	No	6
<i>Social Service Review</i>	1.421	No	No	1

Reporting Guidance Specific to RCTs in “Instructions to Authors”: whether the “Instructions to Authors” section of a journal provided any guidance or referred to any guidelines on reporting RCTs

Trial Registration Required: whether the journal required RCTs to be registered in a trial registry (e.g., [clinicaltrials.gov](http://clinicaltrials.gov)) prior to publication

Eligible RCTs in 2010: number of RCTs in 2010 that met eligibility criteria for this review

**Table 4. Average Compliance of 309 RCTs with Reporting Standards**

<b>Heading</b>	<b>Item</b>	<b>Clinical Psychology</b>	<b>Criminology</b>	<b>Education</b>	<b>Public Health</b>	<b>Social Work</b>	<b>Total Sample</b>
<i><b>External Validity</b></i>							
10 Items	Participants	54.6%	38.2%	37.9%	56.8%	53.2%	48.6%
7 Items	Timing and Setting	43.1%	46.5%	44.8%	52.9%	55.7%	47.0%
29 Items	Intervention: Total	50.4%	42.8%	52.4%	42.6%	48.3%	48.3%
10 Items	Intervention Implementation: Design	74.1%	69.7%	79.7%	68.6%	80.0%	74.4%
12 Items	Intervention Implementation: Delivery	43.8%	35.5%	44.8%	34.4%	37.9%	40.7%
7 Items	Intervention Implementation: Uptake	27.8%	17.1%	26.5%	19.4%	20.7%	24.0%
26 Items	Control: Total	38.4%	38.0%	46.9%	24.8%	22.1%	36.7%
8 Items	Control Implementation: Design	60.5%	62.1%	70.9%	42.7%	43.1%	58.5%
12 Items	Control Implementation: Delivery	32.3%	31.5%	41.4%	19.6%	16.2%	30.9%
6 Items	Control Implementation: Uptake	21.0%	18.8%	25.8%	11.4%	5.8%	19.0%
2 Items	Programme Differences	29.8%	27.4%	27.0%	19.3%	17.5%	25.6%
4 Items	Outcomes*	67.2%	54.8%	53.7%	57.5%	56.3%	59.1%

<b>Heading</b>	<b>Item</b>	<b>Clinical Psychology</b>	<b>Criminology</b>	<b>Education</b>	<b>Public Health</b>	<b>Social Work</b>	<b>Total Sample</b>
<b><i>External Validity</i></b>							
5 Items	Interpretation	75.6%	58.7%	51.2%	50.3%	63.0%	60.3%
83 Items	Total External Validity	48.4%	42.2%	47.7%	40.2%	41.8%	45.3%
<b><i>Internal Validity</i></b>							
9 Items	Trial Design	58.7%	50.9%	50.3%	49.4%	57.2%	53.3%
4 Items	Random Sequence*	30.1%	11.3%	18.0%	28.9%	28.8%	24.4%
13 Items	Data Analysis*	50.0%	31.8%	36.0%	45.8%	44.6%	42.8%
3 Items	Allocation Concealment*	26.3%	17.2%	3.4%	14.8%	28.3%	16.3%
3 Items	Blinding*	20.2%	4.3%	11.2%	14.8%	18.3%	14.7%
8 Items	Participant Flow*	55.4%	14.5%	20.4%	47.1%	37.5%	38.2%
40 Items	Total Internal Validity	47.0%	27.4%	30.0%	40.5%	41.2%	38.3%
<b><i>Study Details</i></b>							
16 Items	Title and Abstract	40.8%	17.9%	28.4%	45.5%	34.7%	35.6%
5 Items	Protocols and Manuals*	29.9%	11.6%	14.6%	19.7%	27.0%	21.2%
3 Items	Ethical Concerns	78.1%	47.3%	41.9%	77.1%	76.7%	64.3%
24 Items	Total Study Details	43.2%	20.3%	27.2%	44.1%	38.3%	36.2%

<b>Heading</b>	<b>Item</b>	<b>Clinical Psychology</b>	<b>Criminology</b>	<b>Education</b>	<b>Public Health</b>	<b>Social Work</b>	<b>Total Sample</b>
<i>Total Score</i>							
	Total Score for All Standards	47.2%	34.6%	39.5%	40.9%	41.1%	41.9%

Number of RCTs in each discipline: RCTs per discipline: Clinical Psychology—99, Criminology—31, Education—89, Public Health—70, Social Work—20

\*Denotes Cochrane Risk of Bias item

### *3.2.6 Reporting Characteristics: Important Study Details*

The majority of trials included a statement about conflicts of interest (70.9%) and whether the authors developed the intervention under investigation (63.4%). About half of the trials (58.6%) reported ethical considerations for social and psychological interventions regarding study design and execution, or data collection, analysis, and reporting. Few papers included information about protocols for the trial (7.4% of studies) or access to treatment manuals (35.6% of studies).

## **4. Discussion**

### ***4.1 Overall Findings***

This comprehensive study substantiates the need for a new reporting guideline for social and psychological intervention trials. Previous reporting guidelines and reporting quality assessment tools indeed have made useful contributions to the advancement of medical and social research, particularly the CONSORT Statement and its official extensions. However, the results of this study indicate that no current guideline is sufficient in its included standards, development, and dissemination. For example, social and behavioural science guidelines could benefit from including a more diverse group of stakeholders in their collaborations, informing guideline content with evidence of poor reporting of the targeted research methodology, utilising rigorous consensus processes to choose guideline content, publishing a complementary and freely-accessible explanation and elaboration (E&E) document, and employing multi-faceted dissemination strategies.<sup>16</sup> These practices would help these guidelines be both more useful and widely utilised by relevant stakeholder communities.<sup>29,30</sup>

Nevertheless, previous reporting guidelines in the social and behavioural sciences have offered an array of tailored advice pertinent to social and psychological intervention

RCTs. Specifically, 89 reporting items were identified that are either not included in any of the CONSORT guidelines or are important modifications of CONSORT standards. These items indicate that, while CONSORT guidance has been rigorously developed and provides extensive guidance for many types of RCTs, it could benefit from more tailored standards for reporting social and psychological interventions.

Our analysis of RCT reporting quality suggests that publications of social and psychological intervention trials often fail to comply with published reporting standards. Our results are consistent with many previous studies on the quality of social and psychological intervention RCTs in specific disciplines.<sup>7-11</sup> While reporting quality varies across disciplines and particular reporting items, compliance with reporting standards is generally low (on average below 50% per study). Moreover, many reporting standards in the social and behavioural sciences are currently vague and underdeveloped, particularly those related to external validity.<sup>12</sup> This imprecision makes it easier for authors to comply with these standards due to their lack of specificity, leading to reports that still do not describe social and psychological interventions or trial methods sufficiently for critical appraisal, replication, and proper inclusion in reviews and meta-analyses. Reporting guidelines need updated standards that provide more specific recommendations based on recent evidence from intervention research in order to improve the usefulness of these trial reports.<sup>31</sup>

#### ***4.2 Guideline Adherence to Best-Practices***

Reporting guideline development and dissemination should follow robust methodology to be useful and widely utilised by relevant scientific communities.<sup>29</sup> Rigorous methods should be employed at each stage of the guideline development process. Otherwise, reporting standards could inadvertently fill articles with details that are

inconsequential to interpretation, place an unnecessary burden on research authors, inadvertently lead to the exclusion of important items from reporting guideline checklists, and prevent use of the guideline. Though previous reporting guidelines in the social and behavioural sciences have been important advances in trying to improve RCT reporting quality, this review indicates a clear need for a more rigorously developed and widely used guideline specifically for social and psychological intervention RCTs (see Appendix E).

#### *4.2.1 Best-Practices for Preliminary Planning*

The first stage of guideline development involves preliminary planning. Akin to this review, reporting guideline development benefits from evidence of poor reporting of the targeted research methodology, as well as identifying potentially relevant reporting standards from existing guidelines. While social and behavioural science guideline developers did conduct some preliminary searches for previous standards, they did not as consistently report systematic investigation of poor reporting of previous research to assist the guideline development process. Previous standards and evidence of poor reporting can inform guideline developers about areas of reporting that especially need to be addressed. In medicine, guidelines for RCTs and systematic reviews, such as the CONSORT<sup>5</sup> and PRISMA<sup>32</sup> Statements, have had far-reaching impact because they use empirical evidence to identify items that are known to contribute to bias and that authors consistently have reported poorly.

Other preliminary activities include identifying a diverse group of stakeholders to participate in developing the guideline, as well as obtaining funding for guideline development. Expertise of participants in guideline development should reflect important considerations of the particular guidance under consideration. For example, a group of

international journal editors, trialists, and methodologists developed the CONSORT Statement and its extensions.<sup>5</sup> Social and behavioural guideline developers at best reported the involvement of 30 stakeholders at some point during guideline development. More stakeholder groups could have been specifically recruited to help in the various stages of developing these guidelines; such groups include researchers from various scientific organisations, statistical experts, research funders, practitioners and policy-makers, and consumer representatives.<sup>16</sup> In addition, only 3 of 33 participants for the Non-Pharmacological CONSORT Extension were involved in psychosocial intervention research.<sup>33</sup> As this extension most closely addresses social and psychological intervention RCTs of any CONSORT guideline, sparse inclusion of psychosocial researchers in its development is one indication of why CONSORT has been used primarily in medical research. Involvement of diverse stakeholder groups across intervention research disciplines is especially essential for any guideline specifically developed for social and psychological intervention trials. However, this level of stakeholder involvement would likely require greater funding for the project; when reported, social and behavioural guideline developers seemed to obtain funding for their posts at their institutions of employment rather than specifically for developing a guideline.

#### *4.2.2 Best-Practices for Consensus Development*

The development of previous social and behavioural science guidelines also generally did not involve recommended consensus development techniques, such as a Delphi process and formal consensus conference. An inclusive consensus development process, involving prescribed methods for obtaining agreement among guideline developers, is a crucial characteristic of developing a rigorous reporting guideline.<sup>17</sup> Such methods are important to pull together the collective knowledge of the various disciplines

that utilise social and psychological interventions, and most ensures that all essential items for these trials are identified and that the inclusion of non-essential items is averted. The Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations for behavioural change interventions,<sup>26</sup> Transparent Reporting of Evaluations with Non-Randomized Designs (TREND) Statement,<sup>25</sup> and the American Psychological Association (APA) Journal Article Reporting Standards (JARS)<sup>28</sup> did involve consensus meetings and iterative discussions with participants. However, they did not report using a Delphi process to develop a list of potential checklist items for the meeting, nor did they reporting using formal consensus conference methods that counter the negative aspects of group decision making, such as conformity, deferment to authority, and restriction of topics discussed.<sup>34</sup> Consequently, in addition to the standards identified in these guidelines, there may be others that would emerge using rigorous consensus processes. Given the breadth of possible standards applicable to psychosocial research, formal consensus methods would be recommended for a social and psychological intervention guideline.

In addition to publishing a reporting guideline checklist, an infrequently employed best practice is to develop another document that provides an explanation of the rationale and evidence for each item, an elaboration on important details of each item, and examples of reporting each item well.<sup>16</sup> The CONSORT-NPT Extension did provide such an extension document; however, while the remit of and some examples in the NPT Extension include psychotherapy and behavioural therapies, the literature predominantly discussed in the E&E involves technical medical interventions, for example optic nerve decompression surgery, pancreaticoduodenectomy, and stent procedures.<sup>35</sup> Only two social and behavioural science guidelines<sup>21,28</sup> published a complementary explanation and elaboration (E&E) document, yet neither of them are open access. For example, while the

scholarly article introducing JARS is freely available online, the need to pay for and order a physical copy of *Reporting Research in Psychology*<sup>22</sup> could prevent researchers from accessing and using it. This in turn may hamper the impact of JARS, for the detailed explication of each reporting standard (found only in *Reporting Research in Psychology*<sup>22</sup>) is essential to effective adherence, and open access is crucial for locating and updating guidelines with new and modified standards via feedback from stakeholders.

#### *4.2.3 Best-Practices for Publication and Dissemination*

Multi-faceted dissemination strategies constitute the final phase of guideline development and are needed to ensure that guidelines achieve their intended impact. A survey of 37 generic reporting guideline developers stated that guideline dissemination can be improved through more active promotion and implementation of the resultant guidance.<sup>29</sup> For example, one potentially useful strategy is to include journal editors in the development process from the beginning, as a means to develop commitment and improve uptake of the final product.<sup>17</sup> Developers should particularly seek guideline endorsement and clear adherence measures from journals and funding agencies to ensure that authors follow guidelines when submitting articles or grant applications. In addition, the guideline should also be promoted at conferences, distributed through appropriate practice groups, made publicly available and free online for anyone to access, and include a complementary explanation and elaboration (E&E) document that details how to adhere to each standard. Given the importance of a reporting guideline's accessibility to its overall impact, unrestricted online availability is a critical publication strategy used to better facilitate the implementation and uptake of reporting standards. Online open access facilitates the widespread use and appraisal of a reporting guideline by all interested stakeholders, which is essential both to its impact and to its feedback for future updates of the guideline.

Without effective dissemination, a well-developed guideline promises little systematic improvement on reporting quality.

### ***4.3 Reporting Standards for Social and Psychological Intervention Trials***

This study yielded a multiplicity of modified reporting standards that are not found in CONSORT guidelines and are currently poorly reported (see Appendix F). If a new CONSORT Extension were to be developed, certain reporting standards may require particular attention or modification due to the nature of social and psychological intervention RCTs. Most modifications related to the external validity of trials, for the CONSORT Statement gives relatively little attention to this construct compared to internal validity.<sup>36</sup> Nonetheless, modified standards indicate that certain trial features related to both internal and external validity may require greater emphasis in social and psychological intervention trials to allow readers to assess the extent to which the results of a study may be influenced by bias and apply to particular settings and populations.

#### ***4.3.1 Internal Validity***

Several modified standards for social and psychological intervention trials related to internal validity. For example, it may not be possible to blind participants and providers of social and psychological interventions, so standards could address the role of awareness on intervention effectiveness as well as any methods used to compensate for this (e.g., keeping participants unaware of study hypotheses).<sup>21</sup> Reporting standards should also focus on how outcome assessors and data analysts were blinded, particularly because many important outcomes in these trials are subjectively measured. Preferences and acceptability of an intervention can also influence participants' engagement and adherence,<sup>12</sup> so standards may ask authors to report "treatment" preferences for both participants and

providers to allow the reader to assess any such biases regarding subjective outcomes.<sup>37</sup> Moreover, many outcomes in social and psychological intervention trials relate to latent constructs that are difficult to measure, and when measured, may be assessed using multiple sources (e.g., participants, providers) and methods (e.g., self-report, observation). Authors should attend to the validity, reliability, and other important psychometric properties of such measures,<sup>38</sup> particularly any information indicating the quality of latent construct measurement (e.g., inter-rater reliability).<sup>7</sup> These measures can also be analysed in several ways, so authors could explain which statistical procedures were performed and the rationale behind them. Other key statistical issues include estimation problems, anomalous data points, arguments for the causes of missing data, and variance–covariance matrices for multivariate analytic systems.<sup>28</sup> Reporting standards relating to these issues could help readers assess the quality of trial methods and the trustworthiness of any effects attributed to the intervention.

#### *4.3.2 External Validity*

Modified standards related to external validity often highlighted intervention theory of change. Current standards can be followed without providing adequate details about theories underlying social and psychological interventions.<sup>36</sup> One identified guideline asks authors to be “explicit about the causal pathways through which an intervention is expected to act and to measure relevant pathway variables” (p. 347).<sup>12</sup> Namely, standards could ask authors to report the active change techniques used in the intervention, the causal processes targeted by these change techniques, and any methods and analyses (e.g., structural equation modelling, process evaluations) used to test theory of change.<sup>26</sup> Inclusion of logic models—as described in the Medical Research Council (MRC) Framework for Complex Interventions<sup>2</sup>—can help elucidate links in causal chains of

highly complex interventions, which can be tested.

Another area of emphasis was intervention fidelity. Interventions are rarely implemented exactly as designed, and social and psychological interventions may be flexibly implemented, so many identified standards related to authors reporting how interventions were actually delivered by providers and actually received by participants. Such information can include core intervention components and techniques, dosage, provider training and supervision, and contamination.<sup>37</sup> As interventions have different relative effects depending on the nature of control groups, details about design, delivery, and uptake should be provided for all trial arms.<sup>39</sup> If trials included process evaluations, authors would ideally describe how intervention delivery evolved throughout the course of a trial. Providing these details can help systematic reviewers identify sources of heterogeneity when synthesising social and psychological intervention trials.<sup>36</sup>

Trial setting and participants were two areas largely targeted by social and behavioural science reporting guidelines. Interventions found effective in trials are not always easily transferred to the field, and more knowledge is needed to support the broader dissemination and implementation of social and psychological interventions. As such, readers would benefit from knowing as much as possible about the contextual dependence of an intervention: that is, what things need to be in place for an intervention to work. Information about context in RCT reports might include factors that are believed to support, attenuate, or frustrate observed effects.<sup>40</sup> Authors could also discuss any empirical information, or psychological and sociological theories broadly construed, about contextual factors that may influence the feasibility, coverage, acceptability, and effectiveness of an intervention.<sup>12</sup> Several aspects of setting and implementation may be important to consider, such as how the intervention was introduced to the service system and delivering organisation(s); administrative and political support; staff readiness;

organisational resources, policy, and philosophy; and the wider service system structure.<sup>41</sup> Because studies may be vulnerable to confounding from secular, contemporaneous events,<sup>42</sup> authors might also take into account and describe the socio-political and historical context of the trial.<sup>27</sup> This information would be especially helpful for understanding “treatment as usual” at the time of the trial.

Regarding participants, authors should clearly describe the targeted population, clarifying the nature, scope, and severity of the problem in the population that motivated the intervention.<sup>21</sup> Reports should also detail the extent that recruitment differed from usual practice, as trial recruitment may differ from how users normally enrol in interventions.<sup>27</sup> Authors could provide appropriate measures of the sample such as socioeconomic, demographic, and cultural characteristics to evaluate how representative the sample is.<sup>43</sup> This information is essential to improving the knowledge base for effective transfer of research findings to real-world settings, and an update of reporting standards to include this information is overdue.<sup>44</sup>

#### *4.3.3 Study Details*

Other details of the trial not related to internal and external validity are important. For example, modified standards emphasised discussing other relevant research when interpreting trial findings. Reporting related research represents a unique challenge to social and psychological interventions, for the research base for these interventions may not be as broadly established as interventions in other disciplines, such as medicine.<sup>21</sup> Data from other studies can provide more specific information on generalisability and the effect of implementation practices on trial outcomes.<sup>45</sup> It is also not uncommon for authors to report data from RCTs in several publications, each with a different focus. Authors should report whether other documents about the trial exist or are in preparation, with some way

of linking all publications on an RCT so that interested readers can more easily access all of the information about the study.<sup>46</sup>

Particular conflicts of interest can also appear in social and psychological intervention trials, such as whether the authors developed and evaluated the intervention themselves.<sup>7</sup> Authors could be asked to report any personal biases that may have influenced or could have the appearance of influencing the research, perhaps with a description of how they were managed in the conduct of the study.<sup>27</sup> Authors may also describe ethical considerations regarding equipoise, data collection and analysis, and presenting findings in a way that honours consent agreements with participants and representatives of research sites.<sup>27</sup>

Given the plethora of possible reporting standards, a formal consensus development process is needed to ensure that any new guidance for reporting these trials provides only minimal, essential items in its checklist that capture these issues, while saving the intricacies of reporting these details to elaboration documents.

#### ***4.4 Strengths and Limitations of the Current Study***

Several limitations of the current study should be noted when interpreting its findings. Firstly, the purposive sampling of RCTs could be susceptible to selection bias, yielding results that may differ from an evaluation of a random sample. This risk of bias should be mitigated by the size of this sample compared to previous reviews and the systematic selection of articles in journals with higher impact factors, which plausibly are at greater risk of *underestimating* deficiencies in reporting quality, and are likely to represent highly-cited trials.<sup>15,47</sup>

Other limitations of this review's methods include the lack of double-coding RCTs from the public health literature, the lack of coding other published manuscripts that

provide details about the trials assessed in this review, and the focus on only five disciplines that evaluate social and psychological interventions. The inter-rater reliability for the 239 double-coded studies was quite high ( $\kappa = 0.71$ ), though it did not reach newly-developed criteria ( $\kappa = 0.80$ ) for assessing the validity of evaluations of RCTs reporting quality,<sup>6</sup> indicating that the data from the later-coded public health RCTs may have had adjustments if double-coded and resolved through discussion. Moreover, as reporting quality was only assessed for the papers identified in this study, key details that were not reported could possibly be in “sibling documents”, such as a trial protocol. However, this study aimed to assess adherence to standards that might be considered for *every* trial publication, and studies rarely identified other publications about the trial (36.9%), trial protocols (7.4%), or trial registration (6.1%). Furthermore, reviewing studies in other related disciplines could provide greater generalisability, particularly RCTs of those social and psychological interventions that target units other than the individual (e.g., organisations, physical places). Nonetheless, this study represents the first systematic examination of reporting quality across various disciplines that research social and psychological interventions.

Estimates of reporting quality were also limited by the rigour of current standards, particularly for the introduction and discussion sections, which are relatively lacking in clarity, making it easier to comply with these standards. In comparison, researchers have identified clear, important aspects of trials related to intervention implementation<sup>41</sup> and external validity<sup>42</sup> that have not been incorporated into prominent reporting guidelines. Our own difficulty in applying some standards reaffirmed the need to develop clear, specific recommendations for social and psychological intervention trials based on best current evidence.<sup>31</sup>

Despite these difficulties, deficiencies in trial reports are both real and important.

Included trial reports were those likely to be cited (i.e., those published in high impact journals) and potentially of better quality than articles published in low impact journals.<sup>15,47</sup> As such, the reports assessed are probably representative of the best trial research in these disciplines. It is also worth noting that, while recommended techniques in guideline development were identified and coalesced by members of the CONSORT Group, previous CONSORT guidelines haven't adhered to all of these practices because CONSORT guidelines preceded and contributed to the "best-practices" statement,<sup>16</sup> as did many of the other guidelines in the social and behavioural sciences. Indeed, this is an emerging field. Nonetheless, this study indicates that current guidelines need to be updated in light of the compilation of recommended techniques for reporting guideline development.

#### ***4.5 Implications***

While these findings indicate several limitations in current guidance for and the quality of reporting social and psychological intervention trials, this study also demonstrates that a unified set of standards can be applied to social and psychological intervention trials. As current CONSORT standards require modifications for social and psychological intervention trials, simply disseminating CONSORT as it stands to social and behavioural scientists would likely be insufficient to markedly improve reporting by researchers in these disciplines.<sup>12,13,30</sup> Useful guidelines have been developed that are more specific to social and psychological interventions than CONSORT, such as the American Psychological Association's Journal Article Reporting Standards (JARS) Statement<sup>28</sup> and the WIDER Recommendations for reporting the content of behaviour change interventions.<sup>26</sup> However, these guidelines have not been developed using as rigorous methods nor are they as user-friendly as CONSORT. They also have not been used as

widely as CONSORT, and they have items that are either too focused (WIDER) or too broad (JARS) for the area of social and psychological intervention RCTs.<sup>26-28,37</sup> Given the prominence of CONSORT internationally and the rigorous development and dissemination practices of the CONSORT Group, an official CONSORT extension seems the best option for any guideline that seeks to facilitate better reporting of the details of these trials.

The CONSORT Statement has been previously extended and modified in the past, and the CONSORT Group welcomes further extensions.<sup>48</sup> Extending the CONSORT Statement to social and psychological interventions could address general challenges of RCTs previously targeted by CONSORT guidance while also incorporating items from other reporting guidelines that are particularly important to psychosocial research.<sup>10,13,39,49</sup> A CONSORT extension would serve as an appropriate culmination of previous work on reporting guidelines related to social and psychological interventions.

A new CONSORT extension for social and psychological interventions has the potential to benefit this area of research in several ways. Developed and disseminated according to recommended techniques, it would serve as an appropriate culmination of previous work on reporting guidelines related to social and psychological interventions by consolidating the theoretical and methodological knowledge base of academics who develop, evaluate, and disseminate social and psychological interventions. Authors would produce more impactful research. Editors and research funders can better identify high-quality studies to publish or fund. Systematic reviewers could more properly appraise included trials. The service and policy field would receive higher quality information. The extended guideline could also serve as a pedagogical tool to teach students about design, evaluation, reporting, and critical appraisal of trials—assisting in training more highly-skilled researchers and enhancing the research capacity of academic departments and centres. Taken together, the benefits of the development and utilisation of this guideline

would help advance the “industry standard” of social and psychological intervention trials.

## **5. Conclusion**

Social and psychological interventions address serious health, social, and related outcomes, and are often used with vulnerable populations. Poor reporting limits the value of rigorous research to improve outcomes for these populations.<sup>2</sup> Advancement in the science of reporting social and psychological intervention trials is important for the improvement of programmes, services, and policies. Better transparency in reporting trials is needed so that judgments can be made by users of evidence about the validity and appropriate applications of research findings.<sup>28</sup> A CONSORT extension alone would not remedy all the inadequacies of how RCTs are designed, executed, and reported. Nonetheless, it would provide guidance for improving the scientific rigour in this field, serve as a foundation for other methods of evaluating social and psychological interventions, and allow stakeholders to improve the use of these trials to inform practice and policy decisions.

## References

1. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.
2. Medical Research Council. *Developing and evaluating complex interventions: New guidance*. London: Medical Research Council; 2008.
3. Pawson R, Greenhalgh T, Harvey G, Walshe K. *Realist synthesis: An introduction*. University of Manchester: ESRC Research Methods Programme; 2004.
4. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: Reporting guidelines and the EQUATOR Network. *BMC Medicine*. 2010;8:24.
5. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698-702.
6. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews*. 2012;11:MR000030.
7. Perry AE, Weisburd D, Hewitt C. Are criminologists describing randomized controlled trials in ways that allow us to assess them? Findings from a sample of crime and justice trials. *Journal of Experimental Criminology*. 2010;6:245-262.
8. Han C, Kwak K, Marks DM, et al. The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. *Contemporary Clinical Trials*. 2009;30:116-122.
9. Naleppa MJ, Cagle JG. Treatment fidelity in social work intervention research: A

- review of published studies. *Research on Social Work Practice*. 2010;20:674-681.
10. Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. A comparison of RCTs in health and education. *British Educational Research Journal*. 2005;31(6):761–785.
  11. Stinson JN, McGrath PJ, Yamada JT. Clinical trials in the Journal of Pediatric Psychology: Applying the CONSORT statement. *Journal of Pediatric Psychology*. 2003;28:159-167.
  12. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: Suggested framework and systematic review. *BMJ*. 2006;333:346-349.
  13. Armstrong R, Waters E, Moore L, et al. Improving the reporting of public health intervention research: Advancing TREND and CONSORT. *Journal of Public Health*. 2008;30(1):103-109.
  14. Moore L, Moore GF. Public health evaluation: Which designs work, for whom and under what circumstances? *Journal of Epidemiology & Community Health*. 2011;65:596-597.
  15. Bennett C, Khangura S, Brehaut JC, et al. Reporting guidelines for survey research: An analysis of published guidance and reporting practices. *PLoS Medicine*. 2011;8(8):e1001069.
  16. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
  17. Moher D, Weeks L, Ocampo M, et al. Describing reporting guidelines for health research: A systematic review. *Journal of Clinical Epidemiology*. 2011;64(7):718-742.
  18. Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time

- but raised questions about how quality is assessed. *Journal of Clinical Epidemiology*. 2011;64:136-144.
19. van Aalst J. Using google scholar to estimate the impact of journal articles in education. *Educational Researcher*. 2010;39:387-400.
  20. Moberg-Mogren E, Nelson DL. Evaluating the quality of reporting occupational therapy randomized controlled trials by expanding the CONSORT criteria. *The American Journal of Occupational Therapy*. 2006;60:226-235.
  21. Nelson DL, Mathiowetz V. Randomized controlled trials to investigate occupational therapy research questions. *The American Journal of Occupational Therapy*. 2004;58(1):24-34.
  22. Cooper H. *Reporting research in psychology: How to meet journal article reporting standards*. Washington, DC: APA; 2011.
  23. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. *Family Medicine*. 2005;37(5):360-363.
  24. Higgins J, Green S, eds. *Cochrane handbook for systematic reviews of interventions* Version 5.1.0 5.1.0 [updated March 2011]. Chichester, UK: Wiley-Blackwell; 2011.
  25. Des Jarlais DC, Lyles C, Crepaz N, the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*. 2004;94(3):361-366.
  26. Abraham C. *WIDER recommendations to improve reporting of the content of behaviour change interventions*. [PDF document]. Available from: <http://www.equator-network.org/reporting-guidelines/wider-recommendations-for-reporting-of-behaviour-change-interventions/>. 2009.
  27. American Educational Research Association. Standards for reporting on empirical

- social science research in AERA publications. *Educational Researcher*. 2006;35(6):33-40.
28. APA JARS Group. Reporting standards for research in psychology: Why do we need them? what might they be? *American Psychologist*. 2008;63:839-851.
  29. Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR Network's Survey of Guideline Authors. *PLoS Medicine*. 2008;5:869-874.
  30. Michie S, Abraham C, Eccles MP, Francis JJ, Hardeman W, Johnston M. Strengthening evaluation and implementation by specifying components of behaviour change interventions: A study protocol. *Implementation Science*. 2011;6:10.
  31. Glasziou P, Chalmers I, Altman DG, et al. Taking healthcare interventions from trial to practice. *BMJ*. 2010;341:384-387.
  32. Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*. 2009 6(7):e1000097.
  33. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
  34. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998;2(3).
  35. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT Group. Extending the CONSORT Statement to randomized trials of nonpharmacologic

- treatment: Explanation and Elaboration. *Annals of Internal Medicine*. 2008;148:295-309.
36. Shepperd S, Lewin S, Straus S, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Medicine*. 2009;6(8):31000086.
37. Davidson KW, Goldstein M, Kaplan RM, et al. Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine*. 2003;26(3):161-171.
38. Fraser MW, Galinsky MJ, Richman JM, Day SH. *Intervention research: Developing social programs*. New York, NY: Oxford University Press; 2009.
39. McGrath PJ, Stinson J, Davidson K. Commentary: The Journal of Pediatric Psychology should adopt the CONSORT statement as a way of improving the evidence base in pediatric psychology. *Journal of Pediatric Psychology*. 2003;28:169-171.
40. Moore L. Research design for the rigorous evaluation of complex educational interventions: Lessons from health services research. *Building Research Capacity*. 2002;1:4-5.
41. Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F. *Implementation research: A synthesis of the literature*. Tampa, FL: National Implementation Research Network, University of South Florida; 2005.
42. Bonell CP, Hargreaves J, Cousens S, et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. *Journal of Epidemiological Community Health*. 2011;65:582-587.
43. Eisenstein EL, Lobach DF, Montgomery P, Kawamoto K, Anstrom KJ. Evaluating implementation fidelity in health information technology interventions. *AMIA Symposium*. 2007.

44. Egan M, Bambra C, Petticrew M, Whitehead M. Reviewing evidence on complex social interventions: Appraising implementation in systematic reviews of the health effects of organisational-level workplace interventions. *Journal of Epidemiological Community Health*. 2009;63:4-11.
45. Durlak JA, DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*. 2008;41:327-350.
46. Perry AE, Johnson M. Applying the Consolidated Standards of Reporting Trials (CONSORT) to studies of mental health provision for juvenile offenders: A research note. *Journal of Experimental Criminology*. 2008;4:165-185.
47. Altman DG. Endorsement of the CONSORT statement by high impact medical journals: A survey of instructions for authors. *BMJ*. 2005;330:1056-1057.
48. Moher D, Altman DG, Schulz KF, Elbourne DR. Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *CMAJ*. 2004;171(4):349-350.
49. Boutron I, John P, Torgerson DJ. Reporting methodological items in randomized experiments in political science. *The ANNALS of the American Academy of Political and Social Science*. 2010;628(112):112-131.

## Chapter 4:

### Identifying Reporting Standards for Social and Psychological Intervention Trials: An Online, International Delphi Process

#### Abstract

**Background:** The literature reviews in the previous thesis chapter have indicated deficiencies in reporting randomised controlled trials (RCTs) of social and psychological interventions. They also indicated that existing reporting guidelines for social and psychological intervention RCTs have important limitations in their content, development, and dissemination, indicating the need for a new guideline in this area. However, prior to any meetings to decide guideline content, experts in the development of research reporting guidelines recommend a consultation survey with a wider group of relevant stakeholders than can be invited to the main guideline development meeting. As the second phase of a project to develop a reporting guideline following this prescribed methodology, the objective of this study was to conduct a modified Delphi process to develop and refine a prioritised list of reporting items to consider for a new guideline for social and psychological intervention trials. A secondary objective was to engage a wide group of stakeholders internationally at an early stage in the CONSORT-SPI project.

**Methods:** Delphi participants were identified using a multistep, iterative approach and met pre-specified eligibility criteria as researchers, practitioners, policy-makers, journal editors, researcher funders, and/or representatives of intervention service users. Participants were asked to rate concepts, identified from the review of reporting standards in the last chapter, for importance in a guideline for reporting social and psychological intervention trials. In Round 1, participants ranked the degree of importance of 77 proposed guideline items on a

scale of 1 to 10, with higher scores indicating higher importance for social and psychological intervention trials. In Round 2, participants ranked whether remaining items not reaching consensus in Round 1 should be included or excluded in a set of minimum reporting standards for these trials. In both rounds, participants had the opportunity to comment on proposed items and nominate items that may be missing from the surveys. The median, inter-percentile range, and counts of rankings for each item—in addition to participants' comments—were used to measure consensus.

**Results:** Three hundred eighty-four participants recommend 36 of 77 items in Round 1 for inclusion in the CONSORT-SPI checklist, while 41 items did not reach consensus based on median scores, inter-percentile range, or participants' comments to reconceptualise and re-rank items in Round 2. These 41 items were split into 61 items based on participants' comments, and 5 new items were proposed. In Round 2, 321 participants recommend 22 of 66 items should be included in the guideline, whereas 44 items remained indeterminate. Participants' comments largely highlighted the desire for a minimal, user-friendly checklist that only includes standards applicable to all trials in this area.

**Conclusions:** This Delphi process indicated that many reporting standards currently found in CONSORT are important to reporting RCTs of social and psychological interventions. A large panel of experts from diverse disciplines and professions has highlighted which concepts are likely to be considered by relevant stakeholders as essential or non-essential for reporting social and psychological interventions trials. These results, along with previous literature reviews and a subsequent consensus meeting of stakeholders, have led to the development of a new reporting guideline for these trials.

## 1. Introduction

Reports of randomised controlled trials (RCTs) of social and psychological interventions can serve many purposes. Future trialists may wish to use an RCT report to guide the design, conduct, and analysis of replication studies. Systematic reviewers and methodologists use reports to synthesise intervention outcomes and study methods. Practitioners may use trial reports to recommend interventions for their clients and as guidance for delivering interventions. Policy-makers and users of intervention research may use trial reports to inform decision-making and advocacy. Moreover, research funders and journal editors require high-quality reporting of completed trials for publication and peer-review purposes. To be attentive to the different needs of these various audiences, reports must be accurate, comprehensive, and transparent.

Despite a long-standing movement within health sciences to improve trial reporting, many reports of social and psychological intervention trials do not describe key details required to appraise and apply their findings. Discussed in detail in the previous chapter of this thesis, a systematic review of these trials revealed that many standards were poorly reported, and poor reporting significantly hinders the utility and credibility of this area of research.<sup>1</sup> Numerous reporting guidelines have been developed to help improve the quality of reporting research in the health and social sciences. However, the previous chapter also showed in detail that existing reporting guidelines for social and psychological intervention RCTs have important limitations in their content, development, and dissemination.<sup>1</sup> Such limitations have implications for stakeholders preparing, running, writing, or using reports of social and psychological intervention trials. These two reviews provide the rationale to continue with the development of a CONSORT extension for social and psychological interventions.<sup>2-5</sup>

Developers of high-quality reporting guidelines often hold face-to-face meetings

with select groups of experts to decide which items to include in the reporting guideline checklists. While there are several ways to generate a list of items for consideration at such a consensus meeting, leaders in the development of reporting guidelines recommend a consultation survey with a wider group of relevant stakeholders than can be invited to the main guideline development meeting.<sup>6</sup> One method they particularly suggest for engaging a wide group of stakeholders in guideline development is the Delphi process. This group facilitation technique uses an iterative, multistage process to exchange information and to synthesise expert opinion into group consensus.<sup>7</sup> The Delphi process involves a series of sequential questionnaires answered anonymously by a panel of stakeholders with relevant expertise in a given area of interest.<sup>8</sup> Summarised responses from each structured questionnaire round are provided to participants after each round via controlled feedback.

Combined, these features allow participants to consider their previous opinions, challenge received ideas, and stimulate discussion on new concepts without pressures to conform to participants of higher status. It has several advantages compared with focus groups and questionnaires. Firstly, via anonymous responding and controlled feedback, it addresses social-cognitive biases that are more likely to lead to conformity in group discussion. By using successive online surveys and controlled summarised feedback, participants may be more likely to consider group opinions in a non-adversarial manner compared to focus group formats.<sup>8</sup> Moreover, participants in Delphi processes do not interact directly with each other, which may help avoid group dominance by certain individuals and decrease pressures toward group conformity. Nonetheless, the method maintains a plurality of views generated through facilitated group interaction. Participants are still informed by the collective opinions of the group and can identify items that may have been missed or thought unimportant, allowing the opportunity for participants to change their opinions in light of feedback.<sup>7</sup> This method is also a cost-effective way to

involve a large number of international and cross-disciplinary participants.<sup>9</sup> It gathers stakeholder opinion without needing to bring participants physically together.

An online, modified Delphi process is a preferred methodology for prioritising reporting items for a consensus meeting because of its strengths in surveying areas of uncertainty and measuring consensus.<sup>10</sup> Furthermore, it fits within a structured approach to guideline development. Delphi processes are increasingly incorporated as part of the development of rigorous reporting guidelines, used for example in developing the CONSORT extension for reporting abstracts of trials in journal manuscripts and conference presentations,<sup>11</sup> the SPIRIT Guidelines for developing and reporting protocols for trials,<sup>12</sup> and the CHEERS Statement for reporting economic evaluations of interventions.<sup>13</sup> The use of the Delphi process in this project conforms with best-practices in guideline development. The best-practices recommend holding a Delphi process *after* conducting systematic reviews that indicate whether an adequate guideline already exists for a given research method and whether the reporting quality of published research is poor.<sup>1,6</sup> Such reviews can provide potential items for the first round questionnaire of a Delphi process. The current Delphi process builds off the systematic reviews in the previous chapter in just this manner.

## **2. Methods**

The primary objective of this study was to identify which items should be considered for inclusion in the CONSORT-SPI checklist at the face-to-face consensus meeting. Consensus meetings are complex and labour-intensive processes; developing a prioritised list of checklist items for consideration at the meeting would increase the chances of efficient deliberation. The secondary objective was to involve a very large group of stakeholders internationally at an early stage in the CONSORT-SPI project. Early and

widespread stakeholder consultation was intended to overcome the issues related to engagement with and acceptance of CONSORT guidance by social and behavioural scientists that were highlighted in Chapters 1 and 3 of this thesis. Consequently, an online Delphi process was conducted in order to measure and reach consensus for potential items for the CONSORT-SPI checklist, based on the views of an international group of expert stakeholders in social and psychological intervention research. Ethics approval for this study was obtained from the Department Research Ethics Committee (DREC) of the Department of Social and Intervention, University of Oxford (Ref: 2011-12\_83).

### ***2.1 Selection of Participants***

Delphi processes are often criticised for their lack of setting explicit eligibility criteria for expert stakeholders serving as participants, as well as for using recruitment strategies that tend to sample only from the Delphi team's professional network.<sup>14</sup> The DPhil candidate, under the guidance of his supervisors, therefore set to establish clear eligibility criteria and a wide-ranging recruitment strategy for this Delphi process.

To enhance credibility and ensure widespread acceptance, invited expert participants represented the key stakeholders of social and psychological intervention trials across professions and relevant disciplines.<sup>7</sup> As such, the recruitment strategy targeted informed and interested participants representing one or more groups of stakeholders that the guideline is intended to influence.<sup>3</sup> Researchers had to publish at least one manuscript of an intervention trial, systematic review, or methodological or statistical procedure. To be eligible as a journal editor, participants had to serve on the editorial board of a journal that has the remit to publish social and psychological intervention trials. To be eligible as a practitioner, participants had to provide services in education, police work, social work, mental health work, community health work, or related positions. To be eligible as a

funder, participants had to hold a position as a funder of social and behavioural science research. To be eligible as a policy-maker, participants had to hold a position as a civil servant, elected official, consultant, or related posts. To be eligible as a consumer group representative, participants had to serve a consumer group organisation as a staff member, volunteer, board of directors member, or related position. As the Delphi questionnaires were written in English, all participants also had to be able to communicate effectively in English.

Potential participants were identified using a multistep, iterative approach similar to those for other Delphi processes to develop reporting guidelines.<sup>12</sup> A 21-person Project Executive and International Advisory Group (IAG) of the CONSORT-SPI project<sup>3</sup> identified an initial list of stakeholders who extensively publish, fund, or utilise social and psychological intervention research. Presentations about the CONSORT-SPI project at relevant conferences in the year before the Delphi process also led to a list of interested potential participants. The social and psychological intervention literature, conference proceedings, and member lists of research societies and organisations were also searched to identify experts in this area of research. Importantly, as a key enabler of the implementation of reporting guidelines is their inclusion in instructions to authors and journal policies, editors of journals listed as “Criminology and Penology”, “Clinical Psychology”, “Education”, “Public Health”, and “Social Work” in the ISI Web of Knowledge 2011 Journal Citation Reports (JCR) for Social Sciences were invited to participate. Due to the significant resources required to check whether all of these journals published social and psychological RCTs, editors from all journals were invited so as not to miss those editors of relevant journals. Given that the journal editors recruitment strategy was likely to identify many people not interested in *intervention trials* specifically, low response rates were expected for those invitees identified in this manner.

To identify those outside of project team's professional network, a commentary written by the Project Executive and co-published in several journals invited other stakeholders to contact the project team or visit the project website to participate in the Delphi process.<sup>2,4,5,15-19</sup> In addition, a "snowball recruitment" approach was used, as participants identified through any of these strategies were also asked to nominate further Delphi participants in the initial invite emails to participate in the Delphi process.<sup>8,20</sup> The goal was to recruit at least 100 participants. A secondary goal was to have at least 25% of participants actively serving on editorial boards of relevant journals, which is why an inclusive strategy was used to identify journal editors; it was deemed more important to wade through many potential participants to identify those few who were eligible and might otherwise be overlooked.

## ***2.2 Selection of Preliminary Items***

The CONSORT-SPI IAG held a virtual meeting three months before Round 1 of the Delphi process to nominate items—identified from the review of reporting standards in the previous dissertation chapter—for the initial questionnaire and to suggest credible participants.<sup>21</sup> Prior to the meeting, the IAG received a draft questionnaire based on literature reviews regarding previous reporting guidelines for social and psychological intervention RCTs and their reporting quality,<sup>1</sup> together with feedback from a consultation held at the 2012 Cochrane Colloquium.<sup>22</sup> Members of the IAG also received copies of the literature review<sup>1</sup> and the published commentary about the project launch.<sup>2,4,5,19</sup>

Feedback from this meeting was then used to refine and finalise the Delphi Round 1 survey, which contained 77 items grouped under the traditional headings of a journal article (Title/Abstract, Introduction, Methods, Results, Discussion, and Other Information). Wording for each item was kept similar to existing guidelines,<sup>12</sup> though slightly adapted

where original wording came from biomedical disciplines with terminology that might be seen as foreign or inappropriate by this study's participants.

## ***2.3 Round 1 Delphi Survey***

### *2.3.1 Procedures*

Procedures for data collection, data analysis, and cut-offs for consensus<sup>7</sup> were decided in light of recommended techniques for guideline development,<sup>6</sup> and previous Delphi processes used to develop reporting guidelines.<sup>11,23</sup>

All correspondence with participants was via email. Participants were sent invitation emails the day that the survey opened, which contained information about the objectives of the CONSORT-SPI project and the Delphi survey. Where relevant, those approached were also asked to provide reasons for declining to participate. Participant confidentiality of responses was ensured, and individual responses were known only to the moderators of the Delphi process (i.e., the DPhil candidate and his supervisors, Paul Montgomery and Evan Mayo-Wilson).

The Round 1 Delphi survey was originally intended to be kept "active" for approximately 4 weeks, although this was extended for 9 weeks due to delays in participant responses, particularly amongst some US government-employed participants unable to respond to email during a government shutdown in the fall of 2013. Reminder emails were sent approximately 10 days before the survey close date to all participants who had yet to decline participation. A further reminder email was sent approximately 3 days before the survey close date to all participants who expressed interest in participating or were recommended by the project's International Advisory Group but had yet to complete the survey.

### 2.3.2 Materials

The Round 1 survey began with a brief introductory page that provided a link to full instructions and an informed consent website, which gave a summary of the objective of the survey and defined consensus for the round. Participants were asked to rate each candidate item on a 10-point Likert scale for their importance to include in CONSORT-SPI: a reporting guideline for social and psychological intervention trials. A rating of 1 corresponded to “not at all important” to include in the guideline (i.e., not important enough to include in the CONSORT-SPI checklist), and a rating of 10 corresponded to “very important” to include in the guideline (i.e., essential for all reports of social and psychological intervention RCTs). Participants were asked to rate the importance of *concepts* underlying each item rather than an item’s specific wording, as the phrasing of items would be decided at a later stage. They were also informed that items with middle or inconsistent rankings would be discussed again in later Delphi rounds. Any items from the CONSORT 2010 Statement that were relevant to a particular section of the questionnaire were provided to participants via a web-link in each respective section of the questionnaire. A list of items for the Round 1 survey mapped out onto a draft CONSORT-SPI checklist can be found in Appendix M.

For each candidate item, participants were also provided an optional free-text comment box to clarify their views if desired. These comment boxes were limited to 1,200 characters in order to manage the amount of qualitative data to be analysed in the short timeframe between Rounds 1 and 2. In addition, participants had an opportunity at the end of the survey to clarify any remaining views and to suggest for future consideration any items that were not proposed in the Round 1 questionnaire. These final comment boxes had unrestricted character count limits. The Round 1 survey also collected demographic information on professional affiliation, age, gender, professional areas

(academic/researcher, practitioner, journal editor, research funder, policy-maker, and consumer group representative), and experience in their professional area(s). Please see Appendix I for a copy of the Round 1 questionnaire.

### 2.3.3 Data Analysis

The median, inter-percentile range, and counts of rankings were calculated for each item and—in addition to participants' comments—were used to assess consensus. To measure average scores and variance, medians and inter-percentile ranges were used rather than means and standard deviations because of the skewed, asymmetrical, and non-normal nature of the data. High-ranking items (i.e., those with medians of 8-10), that had little dispersion, and that did not have significant issues noted in free-text comments were considered to reach consensus for proposed inclusion in the checklist during consensus meeting discussions. Low-ranking items (i.e., those with medians of 1-4) with little dispersion and no significant issues noted in free-text comments were considered to reach consensus for proposed exclusion from the checklist during consensus meeting discussions. Items with middle-rankings (i.e., those with medians of 5-7), wide dispersion of scores (i.e., the inter-percentile range was more than the inter-percentile range adjusted for asymmetry), and/or significant issues noted in the free-text comments were included in the Round 2 questionnaire. Cut-offs for high-ranking, middle-ranking, and low-ranking scores were established *a priori* based on cut-offs used in a previous Delphi process to develop a CONSORT extension<sup>24</sup> that match traditional definitions for indicating importance of items in consensus surveys.<sup>25</sup>

To measure dispersion in Round 1, the Disagreement Index (DI) based on the RAND/UCLA Appropriateness Method was used.<sup>25</sup> This method measures dispersion by comparing the inter-percentile range (IPR) (30% and 70%) to the inter-percentile range

adjusted for symmetry (IPRAS), which provides a continuous measure of dispersion that can easily be adjusted to provide stricter or more relaxed definitions of agreement.<sup>25</sup> The Disagreement Index is the ratio of the IPR to the IPRAS,<sup>26</sup> and typically a DI greater than 1 is considered disagreement.<sup>27</sup> However, as none of the items in Round 1 of this study had a DI above 1, a lower DI of .3 or more was used to indicate dispersion given the desire for a strict level of consensus. Furthermore, categorising those items with a DI level of .3 or more as “having disagreement” was consistent in this dataset with categorising as having disagreement those items that had a lower IPR of 7, which is the cut-off point for medians to be classified as middle- rather than high-ranking.

The DPhil candidate used a framework approach to sort, categorise, and interpret the qualitative data arising from text responses.<sup>28</sup> This approach begins with thoroughly reading and re-reading the material in order for the analyst to be familiarised with and obtain an overview of the data. Subsequently, line-by-line coding was done for all responses, relating the raw data to categories that were grounded in the data itself and that emerged through constant comparison and refinement during coding.<sup>29,30</sup> A thematic analysis was then undertaken by systematically indexing the codes into preliminary themes, charting these themes, integrating similar themes, and relabeling the final themes as appropriate.<sup>29</sup>

To facilitate this process, codes were first grouped and analysed according to sections of the questionnaire (e.g., themes were derived from all codes relating to the title and abstract, then the introduction, and so forth). Themes were based on saliency to the development of a reporting guideline for social and psychological intervention RCTs, particularly themes integral to identifying which reporting items may be most important to include in the CONSORT-SPI checklist. Higher-order themes from all responses in Delphi Round 1 were then developed by mapping themes from each of the questionnaire sections,

grouping themes according to identified patterns across questionnaire sections, and synthesising these grouped themes into aggregate themes for Delphi Round 1. NVivo 10 was used to facilitate this process by serving as a storage and retrieval system for coding data, re-labelling as required, and retrieving codes for thematic analysis.<sup>31</sup> The data and preliminary themes were reviewed by the applicant's supervisors, Paul Montgomery and Evan Mayo-Wilson, whose views informed final interpretations. This collating and analysis of Round 1 results, and the concurrent preparation of the Round 2 survey, took place over four weeks.

## ***2.4 Round 2 Delphi Survey***

### *2.4.1 Procedures*

All participants who completed the Round 1 survey were sent an invitation email to the Round 2 survey the day that the survey opened. As in Round 1, all correspondence with participants was via email. The Round 2 Delphi survey was kept "active" for approximately 10 weeks, as the extended length of the Round 1 survey led the Round 2 survey to overlap with winter holiday. As in Round 1, reminder emails were sent approximately 10 days and 3 days before the survey close date to all participants from Round 1.

### *2.4.2 Materials*

Participants received summaries of the group's quantitative and qualitative responses from Round 1 as well as their individual responses for Round 1 items to inform their responses in Round 2.<sup>8</sup> The Round 2 questionnaire was organised into three parts: items with consensus from Round 1, items that required re-ranking, and new items proposed from Round 1. The first section listed 36 proposed items to include in the CONSORT-SPI

checklist unless participants voiced strong objections in Round 2. Participants were informed that the order and wording of these items was not fixed but would be discussed at the face-to-face CONSORT-SPI consensus meeting in light of comments from Round 1 and any comments from this round. Participants were advised to review their “Delphi Part 1 Items” attachment, and then provide any remaining comments on this list of items in a comment box with an unrestricted character count limit.

The second section listed candidate items that did not reach consensus in Round 1 for participants to re-rank in Round 2. Some Round 1 items were split or their wording was changed for Round 2 in light of Round 1 comments. Participants were advised to review their "Delphi Part 2 Items" attachment for a full list of the items. This list also contained the medians and inter-quartile ranges for each item, as well as each participant’s individual scores for each item. As before, items were organised into sections according to the IMRAD format. For each section of the questionnaire, participants were also given a link to summaries of relevant comments about the original items for that section from the Round 1 questionnaire to consult.

For each item in a section, participants were asked to re-rate whether the item should be “Included”, “Excluded”, or “Optional/Unsure” for a minimum set of reporting standards for all social and psychological intervention trials. As in previous reporting guideline Delphi processes, a change from a 10-point Likert scale in Round 1 to fewer, anchored points in subsequent rounds was chosen to make it easier for participants to indicate (and the researchers to interpret) which items participants suggest for the checklist.<sup>11,12,27</sup> Participants were again provided free-text comment boxes to elaborate on their views if desired, although there was only one free-text box for each *section*, rather than for each *item* as in Round 1. This structure intended to help manage the amount of qualitative feedback to analyse in time for the consensus meeting (or a Round 3, if needed). To

compensate, the character limit for these comment boxes was increased from 1,200 characters to 2,500 characters.

The third section listed the five new items proposed from Round 1 for participants to rate in Round 2. The format and response options for this section were exactly the same as for items in Section 2. Following this section, participants were provided one final comment box (with an unlimited character count limit) to express any views not included in their Round 1 or 2 comments, as well as any remaining comments they would like to make on the wording or content of items, for discussion at the consensus meeting, or on social and psychological intervention trials in general. Please see Appendix J for a copy of the Round 2 questionnaire.

#### *2.4.3 Data Analysis*

Using *a priori* definitions for consensus, items with  $\geq 80\%$  ratings of “Include” were considered to be recommended for inclusion in the CONSORT-SPI checklist at the consensus meeting. Any items with  $\geq 50\%$  ratings of “Exclude” were considered as recommended for exclusion. Any items with  $\geq 30\%$  ratings of “Unsure” would be re-ranked in Round 3. Anything else would be brought to discussion at the consensus meeting as “indeterminate”. It is worth noting here that no items reached  $\geq 30\%$  ratings of “Unsure” in Round 2; as such, there was no Round 3 questionnaire.

As in Round 1, a framework approach<sup>28</sup> was used to analyse and interpret the text responses in Round 2. Though related to Round 1 codes and themes, Round 2 codes and themes were not based on the analysis from Round 1, but again were grounded in the data obtained in Round 2.<sup>30</sup> Using the same approach as Round 1, line-by-line coding was performed for all Round 2 responses, followed by a thematic analysis according to sections of the Round 2 questionnaire, and then higher-order themes were developed from these

themes.<sup>28,29</sup> NVivo 10 was again used to facilitate this process.<sup>31</sup>

Once themes from Round 2 were identified, a further analysis was undertaken to develop superordinate themes for the entire Delphi process. The “higher-order” themes from Round 1 and Round 2 were charted and rearranged according to similarities and differences, and codes were reviewed once more to identify comments from participants that were related to the Delphi process methods or the CONSORT-SPI guidance in general (rather than specific items for the checklist). These codes and the higher-order themes were then synthesised into the superordinate themes emerging from the entire Delphi process. The data and preliminary themes were once again reviewed by the applicant’s supervisors, Paul Montgomery and Evan Mayo-Wilson, to inform the final interpretations. This collating and analysis of Round 2 results took place over five weeks.

To ensure the reliability and validity of the analysis, the DPhil candidate reviewed the hierarchy of themes after drafting the results section of this manuscript, added quotations as needed, refined the text, and re-read the raw text responses again as a safeguard for the interpretations below to accurately reflect the Delphi responses. This process allowed the DPhil candidate to re-examine data in its original context rather than through de-contextualised computer-assisted qualitative data analysis.<sup>32</sup> Drafts were also reviewed by the student’s supervisors, who had familiarity with the raw text responses. Furthermore, the identified themes for Round 1 (with supporting quotations) were reviewed by all participants in between Rounds 1 and 2 as part of the Delphi process, and the themes from Round 2 were reviewed by the CONSORT-SPI consensus meeting participants, who had participated in the Delphi process.

Themes and summaries of text responses from both Rounds 1 and 2 are presented with quotations to support and enrich the qualitative analysis by facilitating transparent appraisal of some source material and making the DPhil candidate’s thematic

interpretations more evident.<sup>33</sup> Due to the vast amount of qualitative data and desire to keep the narrative below to a manageable length, sample quotations provided are limited to those best summarising the key ideas underlying a theme. Anonymous participant ID numbers (and, for Round 1, a participant's ranking for the item) are provided next to each quotation. Minority and alternative views are captured in themes and quotations where applicable to demonstrate the variety of views in the Delphi process that may be in contradiction with either other views or the summary quantitative ranking for an item.<sup>34</sup>

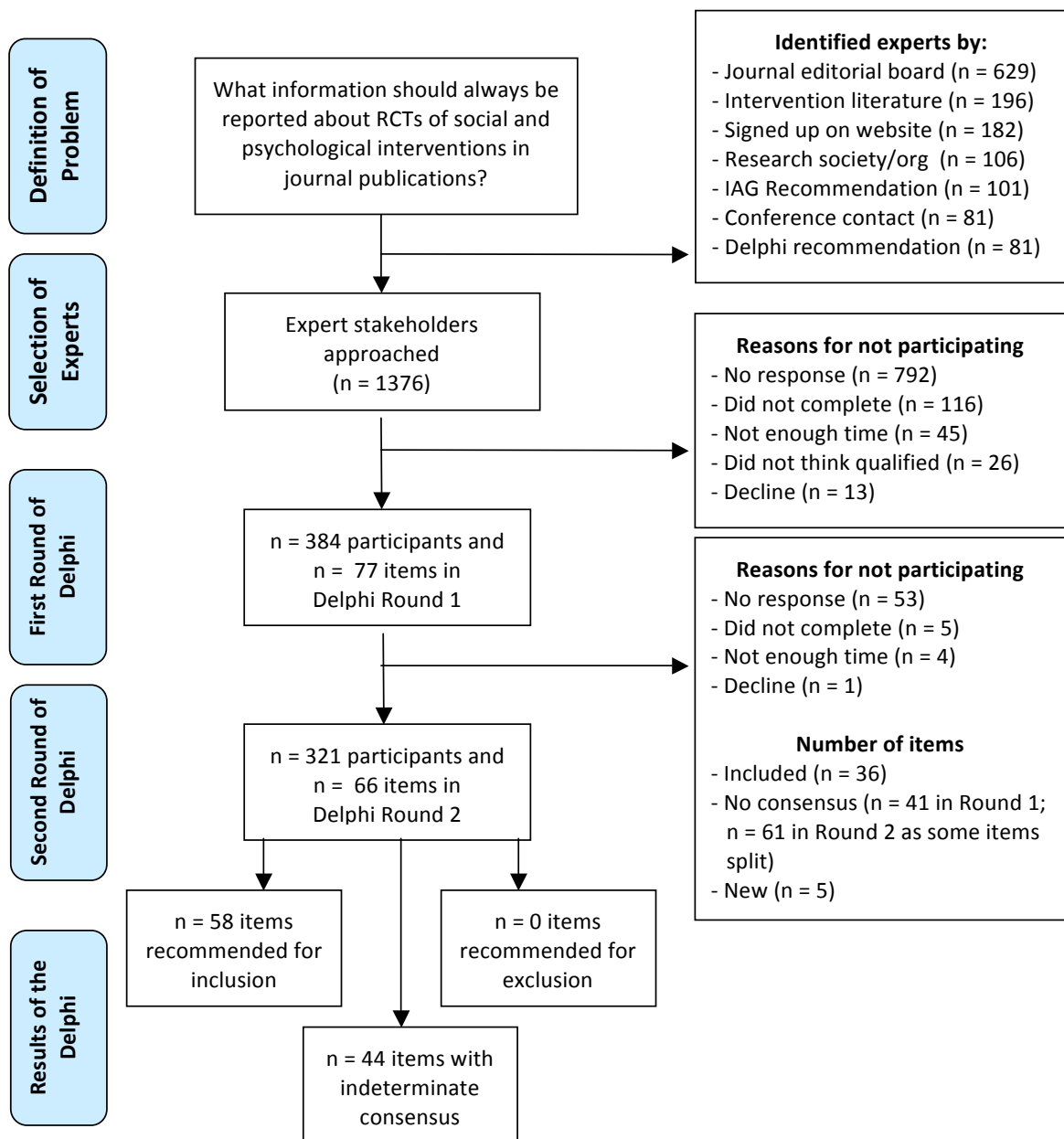
### **3. Results**

#### ***3.1 Delphi Participants***

The flow of Delphi participants can be found in Figure 1. Of the 1376 participants invited to participate in the survey, 584 (43%) responded to the invitation email. Of these, 45 people did not participate citing lack of time (8% of the 584 responding), 26 because they did not think they were qualified (4%), 13 simply declined to participate (2%), and 116 (20%) began the survey but did not complete it.

All 384 participants who agreed to participate and completed the Round 1 survey met pre-specified eligibility criteria for one or more of the professional categories, with 184 female participants (47.9%) and 194 male participants (50.5%); 6 participants (1.6%) did not report gender. Participants represent a wide variety of age groups, reside in 32 countries, and work in over a dozen specialty areas of social and psychological intervention (Table 1). Of the 384 participants, 355 (92%) identified as an academic or researcher, 110 (29%) identified as a practitioner or provider of social and psychological interventions, 132 (34%) are editors of journals in this area, 47 (12%) hold positions involved with funding research, 36 (9%) are involved in policy-making, and 21 (6%) represent recipients of social and psychological interventions (Table 2). Researchers

Figure 1. CONSORT-SPI Delphi Participant Flow Diagram



participating in Round 1 have published a variety of types and amounts of manuscripts related to social and psychological interventions (Table 3). Practitioners, research funders, policy-makers, and consumer representatives demonstrated a range of different career stages (Table 4).

**Table 1. Gender, Age, Country, and Discipline of Delphi Participants (N = 384)**

<b>Demographic</b>	<b>N</b>	<b>Percent</b>	<b>Demographic</b>	<b>N</b>	<b>Percent</b>
<b><i>Gender</i></b>			<b><i>Discipline</i></b>		
Female	184	47.9%	Business	4	1.0%
Male	194	50.5%	Criminology	23	6.0%
Did not report	6	1.6%	Economics	14	3.6%
<b><i>Age</i></b>			Education	46	12.0%
< 35	54	14.1%	Epidemiology	30	7.8%
35 - 44	92	24.0%	Health	27	7.0%
45 - 54	92	24.0%	Services		
55 - 64	107	27.9%	Methodology	29	7.6%
65+	28	7.3%	Nursing	12	3.1%
Did not report	11	2.9%	Policy Studies	17	4.4%
<b><i>Country</i></b>			Psychiatry	17	4.4%
Australia	17	4.4%	Psychology	135	35.2%
Austria	2	0.5%	Public Health	51	13.3%
Bangladesh	1	0.3%	Social Work	20	5.2%
Belgium	2	0.5%	Sociology	22	5.7%
Brazil	3	0.8%	Statistics	18	4.7%
Canada	14	3.6%	Various Health	33	8.6%
Chile	1	0.3%	Sciences		
China	4	1.0%	Did not report	48	12.5%
Colombia	2	0.5%	Note: Percentages add up to more than 100%, as some participants indicated expertise in more than one discipline		
Denmark	8	2.1%			
Finland	1	0.3%			
France	1	0.3%			
Germany	9	2.3%			
Greece	2	0.5%			
India	1	0.3%			
Iran	1	0.3%			
Ireland	3	0.8%			
Israel	1	0.3%			
Italy	3	0.8%			
Japan	7	1.8%			
Malaysia	1	0.3%			
Netherlands	8	2.1%			
Norway	6	1.6%			
Poland	1	0.3%			
Portugal	2	0.5%			
South Africa	2	0.5%			
South Korea	1	0.3%			
Spain	3	0.8%			
Sweden	3	0.8%			
Switzerland	2	0.5%			
UK	125	32.6%			
USA	147	38.3%			

**Table 2. Professional Roles of Delphi Participants**

<b>Professional Role</b>	<b>N</b>	<b>Percent of Sample</b>
<b><i>Academic/Researcher</i></b>	<b>355</b>	<b>92.4%</b>
Trialist	178	46.4%
Systematic Reviewer	211	54.9%
Statistician	45	11.7%
Methodologist	155	40.4%
Other	75	19.5%
<b><i>Practitioner</i></b>	<b>110</b>	<b>28.6%</b>
Teacher/School Administrator	22	5.7%
Police Worker	3	0.8%
Social Worker	18	4.7%
Mental Health Worker	38	9.9%
Community Health Worker	6	1.6%
Other	47	12.2%
<b><i>Journal Editor</i></b>	<b>132</b>	<b>34.4%</b>
Criminology Journal	7	1.8%
Education Journal	19	4.9%
Psychology Journal	38	9.9%
Public Health Journal	21	5.5%
Social Work Journal	10	2.6%
Other	54	14.1%
<b><i>Research Funder</i></b>	<b>47</b>	<b>12.2%</b>
Government	31	8.1%
Non-Profit	14	3.6%
Commerical	4	1.0%
Other	7	1.8%
<b><i>Policy-Maker</i></b>	<b>36</b>	<b>9.4%</b>
Civil Servant	11	2.9%
Elected Official	0	0.0%
Policy Consulting	13	3.4%
Other	13	3.4%
<b><i>Consumer Group Representation</i></b>	<b>21</b>	<b>5.7%</b>
Staff for Consumer Advocacy Group/Organisation	1	0.3%
Volunteer for Consumer Advocacy Group/Organisation	5	1.3%
Board of Directors for Consumer Advocacy Group/Organisation	12	3.1%
Other	4	1.0%

**Table 3. Number of Intervention-Related Publications by Academics/Researchers in Delphi Process**

Type of Publication	Number of Publications					Total N	Percent of Sample
	0	1	2 to 5	6 to 10	10+		
Trials	19 (10.7%)	22 (12.4%)	69 (38.8%)	37 (20.8%)	31 (17.4%)	178	46.4%
Systematic Reviews	18 (8.7%)	23 (11.1%)	79 (38.2%)	40 (19.3%)	47 (22.7%)	207	53.9%
Statistical Publications	10 (22.2%)	4 (8.9%)	19 (42.2%)	2 (4.4%)	10 (22.2%)	45	11.7%
Methodological Publications	24 (15.6%)	19 (12.3%)	52 (33.8%)	26 (16.9%)	33 (21.4%)	154	40.1%
Other Intervention Publications	8 (11.1%)	3 (4.2%)	22 (30.6%)	10 (13.9%)	29 (40.3%)	72	18.8%

NB: The same researcher could indicate having published multiple types of publication.

Under “Number of Publications”, the number to the left indicates the number of participants who have published that many of that type of publication, and the parenthetical indicates the percentage of participants who published a certain number of a certain type of publication. For example, 178 (46.4%) of the 384 participants in the Delphi process identified as trialists. Of these 178 participants, 19 (10.7%) have not actually published any trials. 22 (12.4%) of trialists have 1 trial publication, 69 (38.8%) have 2-5 trial publications, 37 (20.8%) have 6-10 trial publications, and 31 (17.4%) have 10 or more trial publications.

“Total N” indicates the number of participants in the overall sample who have published that type of publication.

**Table 4. Years of Experience by Professionals in Delphi Process**

<b>Years in Position</b>	<b>N</b>	<b>Percent of Sample</b>	<b>Median Years</b>	<b>Min Years</b>	<b>Max Years</b>
<b><i>Practitioners</i></b>	<b><i>110</i></b>	<b><i>28.6%</i></b>			
Teacher/School Administrator	22	5.7%	11.0	0	37
Police Worker	3	0.8%	3.0	1	27
Social Worker	18	4.7%	24.5	0	40
Mental Health Worker	38	9.9%	23.0	3	41
Community Health Worker	6	1.6%	5.0	2	10
Other	46	12.0%	17.5	1	40
<b><i>Research Funders</i></b>	<b><i>47</i></b>	<b><i>12.2%</i></b>			
Government	28	7.3%	4.0	0	40
Non-Profit	12	3.1%	5.0	0	11
Commercial	4	1.0%	3.5	0	20
Other	6	1.6%	10.0	1	30
<b><i>Policy-Makers</i></b>	<b><i>36</i></b>	<b><i>9.4%</i></b>			
Civil Servant	11	2.9%	24.5	2	33
Elected Official	0	0.0%	N/A	0	0
Policy Consulting	13	3.4%	10.0	1	25
Other	12	3.1%	7.5	0	30
<b><i>Consumer Representative</i></b>	<b><i>21</i></b>	<b><i>5.5%</i></b>			
Staff for Consumer Advocacy Group/Organisation	1	0.3%	0.5	0.5	0.5
Volunteer for Consumer Advocacy Group/Organisation	5	1.3%	3.0	2	15
Board of Directors for Consumer Advocacy Group/Organisation	11	2.9%	6.0	1	30
Other	3	0.8%	0.0	0	4

### **3.2 Round 1 Results**

#### *3.2.1 Quantitative Rankings*

In Round 1, participants recommended 36 of 77 items for inclusion in the CONSORT-SPI checklist (Table 5). The remaining 41 items did not reach consensus based on median scores, inter-percentile range, or participants' comments to reconceptualise and re-rank items in Round 2 (Table 6). These 41 indeterminate items were split into 61 items to be ranked in Round 2 based on participants' comments (see Appendix K for decision rules). To track the progress of the potential CONSORT-SPI checklist, the items from Round 1 mapped onto the CONSORT 2010 checklist are provided in Appendix N.

#### *3.2.2 "Title and Abstract" Text Responses*

For the items on the title and abstract, concern was raised about word length and overly-prescriptive titles. For example, the population and intervention might be important to include in the title, but requiring authors to report primary outcome, control, and time of follow-up might be too long. Such details may be more appropriate for the abstract than the title: "To me titles are less critical than abstracts for conveying these details" (P021: 6).

#### *3.2.3 "Introduction" Text Responses*

Regarding items in the introduction, it may be important to distinguish the experimental intervention from other interventions, yet it is difficult to choose which "other" interventions to discuss: "Often it is not possible to fully describe all interventions for a particular problem and so a selective review might be needed" (P147: 5).

Expanded documentation on interventions in the introduction is important for CONSORT-SPI to address, indicating for example the "maturity" of evidence on the

intervention: “There should be emphasis on systematic reviews.... the authors should state why their trial was necessary” (P152: 10). In particular, authors should indicate how the intervention was hypothesised to work, but the proposed wording of “theory of change” was a problematic way of conceptualising this item: “This section needs to help identify theoretical underpinnings of study (e.g. application of relevant theories to conceptualize the intervention, which goes beyond theory of change)” (P282: 9).

Objectives or hypotheses for all “levels” of an intervention should be specified, rather than focusing on clusters: “if the theory of change involved impacts through a level 'higher' than individuals, then the entire theory in its complexity ... should be specified” (P064: 7). Furthermore, it may be important to justify hypotheses, yet an explicit statement of this is not always needed and overlaps with other proposed items about prior research: “This links back to the review of prior research (on the target problem and intervention/s under study)” (P080: 10).

### 3.2.4 “*Trial Design*” Text Responses

The major theme from this section was that incentives may be important aspects of trials in this area, yet they are too infrequently reported: “This information is strikingly underreported in the literature. Authors would benefit from guidance about the level of detail to present” (P167: 10).

Describing what actually happened in the comparator conditions is also key: “Not only why they were chosen but what happened in the control group. They are not neutral experiences... and different countries have more or fewer services available” (P073: 10).

### 3.2.5 “*Randomisation and Blinding Procedures*” Text Responses

There was some scepticism about the feasibility and importance of blinding data

analysts in this area: “For social science, ... most people do their own analysis, they do not hire someone else, so this is not really so useful” (P066: 1). As such, the availability of trial data may be a more viable and useful alternative: “Important, but perhaps more important to make data files publically available so that analyses can be replicated” (P271: 8). Blinding participants and providers was also noted as often impossible in this area, which should be discussed in the guideline: “For this -- and other blinding issues -- the narrative accompanying the statement should discuss the impossibility of many aspects of blinding in behavioral studies” (P080: 8). Consequently, authors should consider reporting strategies other than blinding to address biases that blinding procedures target: “I think CONSORT should ask for any strategy to minimize 'demand characteristics / contamination type' issues, if blinding was not possible” (P135: 10).

### 3.2.6 “*Methods: Process Evaluation*” Text Responses

Participants voiced mixed views about the proposed item on trial setting. In general, contextual information is very important: “Much of these features are about the context in which interventions are delivered/implemented. Again, this is key because it can provide much explanatory info. the [sic] conditions under which interventions are effective” (P282: 8). However, the item on context may need to be split, as some aspects of context might be more important than others: “not sure all of these things are equally important” (P042: 7). Similarly, items on context could provide more concrete guidance about requested details; otherwise, it will be hard for authors to know what to report: “Strikes me as sort of a rest category; not sure what I would have to fill out here and why” (P135: 7). There was concern that an appropriate understanding of which contextual variables are important does not yet exist: “I don't think we are near being able to understand effects of these variables, some of which are more likely to be second order” (P269: 8).

Similar to the above, methods used to investigate the influence of context on study outcomes may not be part of every study, and there was scepticism about methodological techniques in this area: ““context” is not randomised, so this question encourages the incorrect belief that interaction terms in trials can be interpreted as if they had the same status as main effects - which is not true” (P081: 5). As such, a description of context may be more appropriate than investigations of contextual dependence of outcomes: “A detailed description of the context, as suggested under item 28, is more important and more feasible than a detailed investigation of how context is likely to influence the study outcomes” (P112: 3).

Attention to mediators was identified as a growing area for theory testing, yet such testing may not be done for some causal mechanisms or any at all in some studies: “not sure you can always measure all the causal mechanisms” (P006: 5). A purely quantitative conceptualisation of this mediation testing was also deemed too narrow: “As well as how qualitative data were used to capture unanticipated outcomes and explore complex mechanisms” (P224: 10).

One emerging theme related to the interventions in a trial was that details for implementing the intervention and comparator are important to this area, but “precise details” on these aspects may be too much for the main trial paper: “can refer to separate documents kept online for this as this is not necessary in the main report” (P040: 9).

Moreover, details on methods to assess or enhance implementation fidelity may be important yet are often missing; moreover, the conceptualisation of this concept in the Round 1 item may have been problematic: “Important, although quality of provision will often require qualitative data, so the term measures is problematic” (P224: 10).

Information needed about intervention implementation may also depend on the “phase” of the trial: “Some issues, like this one, are less important to describe in detail for efficacy

trials. Variations in things like this are important for effectiveness trials” (P090: 6).

### 3.2.7 “Methods: Outcomes and Data Analysis” Text Responses

For outcomes and data analysis, imputation was seen as important, yet some participants wanted to rank separately what variables were used for imputation, the number of imputations performed, and the rationale for any transformations, if done: “Also a rationale for why transformations were necessary (e.g. was the statistical test not robust against the violations in the data?)” (P154: 9).

Some participants also wanted to split the item on the measurement of primary and secondary outcome measures, as there were too many concepts in this item: “I agree with the “clearly defined” and how they were measured, but not methods used to enhance” (P033: 7). Moreover, the distinction between primary and secondary outcomes is not as important as reporting all outcomes in this area: “for complex social interventions, is the distinction between primary and secondary outcomes always clear (apart from in relation to sample size calculations)?” (P112: 10).

Another theme was the importance of describing measures and their psychometric properties, though references to measures rather than providing copies of measures may be more feasible: “Often times the documents containing the info are pay per use, not sure this is feasible, could be changed to: References of validation studies of measures [sic] used” (P006: 4).

### 3.2.8 “Results: Participant Flow and Recruitment” Text Responses

When reporting baseline data, one theme was that it is difficult to decide which baseline characteristics are key or theoretically important: “While important, it is hard to define “key” characteristics. I think that some information is absolutely critical, and it

might be helpful to have some guidance on what variables are required at a minimum” (P021: 8). In addition, there was concern that the conceptualisation of the item as it stands in CONSORT is more medical rather than social or behavioural: “clinical characteristics is an inappropriate use of terms for many social and psychological interventions. Very medically inspired” (P120: 8).

Regarding comparisons of study completers with the full sample, it may be best to simply clarify the completeness of data for each analysis and how any missing data were handled, as there was concern about the trustworthiness of statistical analyses that test significant differences between study completers and the sample at baseline: “i think simple comparison of baseline differences can provide misleading reassurance that no differences. also risks over interpretation of spurious differences” (P205: 1).

### 3.2.9 “Results: Process Evaluation” Text Responses

References or brief descriptions of details on intervention implementation may be more feasible than “precise details”: “Maybe - but difficult to compare across. We need a standard language - and a few crude categories would be so much more helpful than a detailed description” (P093: 6). It was also pointed out that acceptability was conceptually different from details of intervention and comparator taken up; consequently, acceptability was assessed on its own as an item in Round 2 even though the Round 1 item received high, homogeneous rankings. Thus, the recommended item became “Precise details of the intervention(s) and comparator(s) actually taken up by participants”, and the rated Round 2 item became “Describe any information about the acceptability or perceived value of the intervention(s) and comparator(s) by participants”. Furthermore, details about intervention roll-out, organisational capacity, and other barriers and facilitators of implementation were thought important but possibly not relevant for all trials: “this would be relevant only in the

case of implementation studies - for an efficacy study, I think not necessary” (P006: 4).

As with related items in the methods section, results of analyses evaluating causal mechanisms and contextual dependence of outcomes were thought to be important yet not relevant or done in all trials, and participants voiced further concerns about the validity of statistical analyses in this area: “these statistical analyses tend to be poorly executed and misleading” (P013: 3).

### *3.2.10 “Results: Outcomes and Estimation” Text Responses*

Regarding outcome data, it may be helpful to somehow connect all of the papers reporting outcomes of a trial, but the item in Round 1 could benefit from re-conceptualisation: “The topic is important, but this statement is way too broad. “Any other analyses performed” is ill defined and far too all encompassing” (P038: 4).

Rather than provide variance-covariance matrices, it was suggested to provide the relevant data to allow for replication, especially as these matrices may not be relevant to all trials: “All data needed for replication should be provided rather than summaries” (P134: 6).

While data on harms is important, a separate item for adverse events may not be needed for the CONSORT-SPI Extension beyond what is already in CONSORT: “This is really just a subset of 52a - and this may vary quite a lot, depending on the intervention. I would propose to add examples of such adverse effects under 52a” (P112: 5).

### *3.2.11 “Discussion” Text Responses*

For the discussion section, generalisability broadly construed is an important area for CONSORT-SPI to target. However, a more beneficial use of the guideline may be to request reporting information about the study sample and setting for readers to then

interpret for their own contexts, rather than requesting lengthy author interpretations:

“Important to emphasize generalizability, but if the sample and context were well-reported in the methods and results, there's not a whole lot more to say here that isn't speculative, other than to reinforce the point that findings only apply to the population and context tested, and to highlight any other studies that have assessed the intervention with other populations or contexts” (P271: 7).

The feasibility of authors to state the implications of trial findings to future research, policy, and practice was also questioned: “In my experience, study investigators are often not well trained to consider policy implications in a systematic way” (P009: 4).

Though it had relatively low quantitative rankings, an item on alternative interpretations of trial results was thought by some to be important in reducing confirmation bias and tying the trial to other intervention research studies: “Yes, challenge people to invest in alternative hypotheses” (P135: 10).

Trial limitations were seen as important, though the item on limitations was asked to be split in Round 2, as it contained many concepts, and some participants commented that the current item was too prescriptive: “Would like to give researchers some freedom in what they report in the discussion” (P120: 6). It was also noted that discussion of the limitations of process evaluations may not be an item that applies to all trials: “This is important but perhaps some of this information is best placed in a separate paper linked to a process evaluation.” (P070: 10)

### 3.2.12 “Other Information” Text Responses

In the section on “Other Information”, trial registration was noted as an important issue, particularly for linking all papers on a trial, yet many if not most trials in this area are not registered: “Trial registry is rare in the social sciences. Of course it should be

reported if it was registered” (P236: 2).

The item on ethical approval might benefit from re-wording (as ethical approval is always needed), yet it was also noted that an item on ethics may not be necessary given journal requirements: “This would be best served to be provided in short summary to the journal to be able to ensure that ethical guidelines were met, but is not a necessity for the actual publication” (P283: 8).

The item on references to other sources of information on trial methods and outcomes also required re-conceptualisation in Round 2: ““All sources" doesn't fit because they can be redundant. What we want is "sources of all relevant information”” (P038: 7).

The role of the intervention developer in the trial was considered a very important potential conflict of interest when existent: “This really is our biggest conflict of interest. Financial conflicts should obviously be stated, but reputation and career advancement need to be recognized far more often” (P109: 10). As this conflict of interest may not be relevant to all trials, one overarching conflicts of interest statement was proposed: “I think one overarching statement about conflicts and how they are managed rather than three separate as in 63, 64, 65” (P127: 5).

### *3.2.13 Additional Items*

In the final comments section, 5 new items were proposed for ranking in Round 2: costs of the intervention, results of power analyses, methods to avoid or minimise contamination or spillover effects in the trial, stakeholder involvement in the trial, and acknowledgments of those who contributed substantially to the project but did not meet authorship requirements.

**Table 5. Items Recommended for Inclusion in the CONSORT-SPI Checklist from Round 1**

<b>Round 1 Item</b>	<b>Round 1 Median (IQR)</b>
<b>Title</b>	
1. Identify as randomised/randomized in the title	10 (8, 10)
<b>Introduction</b>	
3. Scientific background and explanation of rationale of the study	10 (8, 10)
4. Describe the problem(s) or issue(s) that the intervention(s) is intended to address	10 (9, 10)
8. Specific objectives or hypotheses of the study	10 (9, 10)
<b>Methods: Trial Design</b>	
11. Description of trial design (such as cluster, factorial, crossover), including allocation ratio	10 (9, 10)
12. Report all inclusion and exclusion criteria for participants, providers, settings, and (if relevant) clusters	10 (9, 10)
13. Important changes to methods after trial commencement (such as eligibility criteria), with reasons	10 (9, 10)
14. How sample size was determined	9 (8, 10)
15. When applicable, explanation of any interim analyses and stopping guidelines	8 (7, 10)
<b>Methods: Randomisation and Blinding Procedures</b>	
16. Method used to generate the random allocation sequence	9 (7, 10)
17. Type of randomisation (e.g., minimisation, stepped-wedge) and details of any restriction (such as blocking and block size)	10 (8, 10)
18. Mechanism used to implement the random allocation sequence (such as sequentially numbered opaque envelopes), describing any steps taken to conceal the sequence until interventions were assigned	9 (7, 10)
19. Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	9 (7, 10)
20. Whether and how providers and participants were blind after assignment to interventions, and if maintenance of blinding was assessed	9 (8, 10)
21. Whether and how outcome assessors were blind after assignment to interventions, and if maintenance of blinding was assessed	10 (8, 10)
<b>Methods: Process Evaluation</b>	
25a. Precise details of the content of the intervention(s) and comparator(s) as designed for the study, including clear definitions of the essential and non-essential components for all groups, and the intended differences across groups	10 (9, 10)
25b. Precise details of the intended duration and frequency of the intervention(s) and comparator(s)	10 (8, 10)
25c. Precise details of the intended format of the intervention(s) and comparator(s), such as individual vs. group, in-person vs. electronic provision	10 (8, 10)

<b>Round 1 Item</b>	<b>Round 1 Median (IQR)</b>
<b>Methods: Outcomes and Data Analysis</b>	
32. Explain the choice of outcomes, their timing and length of follow-up, and any differences across groups in how outcomes are measured	9 (8, 10)
34. Any changes to trial outcomes after the trial commenced, with reasons	10 (8, 10)
35. Statistical methods used to compare groups for primary and secondary outcomes, with reasons	10 (9, 10)
37. Methods for additional analyses, such as subgroup analyses, adjusted analyses, and how these compare to the trial registration and protocol	9 (8, 10)
38a. Imputation methods for handling missing data, and whether these methods were pre-specified	9 (8, 10)
<b>Results: Participant Flow and Recruitment</b>	
39. A flowchart including the following for each group: the numbers of participants, clusters, and providers or centres who were (1) approached, (2) screened, (3) eligible, (4) randomly assigned, (5) received the intended intervention, and (6) were analysed for the primary outcome, including the number of participants by each provider or center and reasons for dropout	10 (9, 10)
40. For each group, losses and exclusions after randomisation, including the number of participants who discontinued the intervention but remained in the trial, together with reasons	10 (9, 10)
42. Whether the trial has ended or was stopped, with reasons if so	10 (8, 10)
46. For each group, number of participants (denominator) included in each analysis, and whether each analysis was per protocol or based on initial intervention assignment	10 (8, 10)
<b>Results: Process Evaluation</b>	
47a. Precise details of the intervention(s) and comparator(s) actually offered by providers, with reasons for any differences from design	9 (8, 10)
47b. Precise details of any tailoring by providers of the intervention(s) and comparator(s) to individual participants across groups	9 (7, 10)
48a. Precise details of the intervention(s) and comparator(s) actually taken up by participants, including acceptability if assessed	9 (7, 10)
48b. Amount of the intervention(s) and comparator(s) actually received by participants (e.g., sessions attended) across groups	9 (8, 10)
<b>Results: Outcomes and Estimation</b>	
50a. For each quantitative outcome, the results for each group as well as the estimated effect size and its precision (such as 95% confidence interval)	10 (9, 10)
50b. For binary outcomes, presentation of both absolute and relative effect sizes is recommended	9 (8, 10)
52a. All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	10 (8, 10)
<b>Other Information</b>	
62. References to intervention manual(s), websites, and other resources concerning the intervention	9 (7, 10)
63. Sources of funding and other support, and the role of funders in the design, conduct, analysis and reporting of the trial	10 (8, 10)

**Table 6. Indeterminate Items from Round 1**

<b>Round 1 Item</b>	<b>Round 1 Median (IQR)</b>
<b>Title</b>	
2. The title should be structured around an acknowledged question format, e.g., PICOT (Population, Intervention, Control, primary Outcome, Time of follow-up)	7 (5, 9)
<b>Introduction</b>	
5. Describe previous research on the experimental intervention(s)—intervention development, pilot-testing, evaluations, and systematic reviews	9 (8, 10)
6. Describe research about other interventions for this problem or issue	7 (6, 9)
7. Describe the intervention and its hypothesised theory of change	9 (8, 10)
9. Whether any objectives or hypotheses pertain to the cluster level	8 (6, 10)
10. How objectives or hypotheses were derived	7 (5, 8)
<b>Methods: Trial Design</b>	
12a. Report why the particular control/comparator intervention(s) were chosen for the trial	9 (8, 10)
<b>Methods: Randomisation and Blinding Procedures</b>	
22. Whether and how data analysts were blind after assignment to interventions, and if maintenance of blinding was assessed	8 (6, 10)
23. Methods to compensate for lack of blinding at any stage	8 (7, 10)
24. If relevant to issues of blinding, description of the similarity of intervention(s) and comparator(s)	8 (6, 9)
<b>Methods: Process Evaluation</b>	
26. Precise details of methods to assess or enhance implementation fidelity of intervention(s) and comparator(s), including the quality of provision and compliance by participants, with measures used	9 (8, 10)
27. Precise details of the plan for implementing the intervention(s) and comparator(s), such as staff recruitment and selection, staff training and support, and physical or technical resources	8 (7, 10)
28. Describe important features of the setting(s) for data collection and intervention implementation, including date and time of study procedures, geographic location, and characteristics of the implementing organisation(s)	8 (7, 9.25)
29. Describe how causal mechanisms were measured and analysed to assess mediators of the intervention(s)	8 (7, 10)
30. Methods used to investigate context and the influence of context on study outcomes	8 (6, 9)

Round 1 Item	Round 1 Median (IQR)
<b>Methods: Outcomes and Data Analysis</b>	
31. Clearly defined primary and secondary outcome measures, their level of measurement, how they were measured, methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors), and how these compare to the outcomes listed in the trial registration and protocol	10 (9, 10)
33. Copies of measures used and their psychometric properties, or references to publicly available documents containing this information	9 (7, 10)
36. Any transformations to quantitative data, and statistical software used	9 (7, 10)
38b. If done, what variables were used for imputation, and the number of imputations performed	9 (7, 10)
<b>Results: Participant Flow and Recruitment</b>	
41. Dates defining the periods of recruitment and follow-up of individuals and clusters	8 (7, 10)
43. Conditions of consent and incentives provided to participants and/or clusters to enrol in the trial, to use the intervention, or to complete outcome measures	9 (7, 10)
44. All theoretically important variables measured at baseline, with data for key baseline demographic, socioeconomic, and clinical characteristics for each group	10 (8, 10)
45. Describe how the full study sample compares with study completers	9 (8, 10)
<b>Results: Process Evaluation</b>	
47c. Precise details of actual professional qualifications, training to deliver the intervention(s), and supervision of providers across groups	8 (6, 9)
49a. Precise details about the actual implementation process, such as intervention roll-out, organisational capacity, and other barriers and facilitators of implementation	8 (7, 10)
49b. Describe features of the broader context important to intervention implementation and observed outcomes, such as concurrent events, area demographics, and the policy-related environment	7 (6, 9)
49c. Results of analyses evaluating causal mechanisms and contextual dependence of outcomes, with evidence to support any claims	8 (7, 10)
<b>Results: Outcomes and Estimation</b>	
50c. Provide any associated variance-covariance matrices for multivariate analytic systems (e.g., multiple regression, structural equation modeling)	8 (5, 9)
51. Report or provide a reference for results of any other analyses performed, including subgroup and adjusted analyses, distinguishing pre-specified from exploratory	8 (7, 10)
52b. Adverse psychological events and increased social disadvantage, indicating the level at which the harm may have occurred (e.g., individual, family, community)	9 (8, 10)

Round 1 Item	Round 1 Median (IQR)
<b>Discussion</b>	
53. Trial limitations, addressing sources of potential bias, imprecision, clinical heterogeneity, inconsistency in response to intervention, multiplicity of analyses, choice of the comparator, lack of complete blinding, and unequal expertise of providers or organizations	10 (8, 10)
54. Limitations in the collection and analysis of process evaluation data, such as information about the delivery and uptake of interventions, context, and intervention acceptability	9 (7, 10)
55. Generalisability of the study findings to related populations and settings, considering the influence of intervention implementation, choice of comparator, sample characteristics, and data about contextual factors	9 (7, 10)
56. Interpretation consistent with results, considering moderators and mediators, balancing benefits and harms, and discussing other relevant evidence	9 (8, 10)
57. Alternative interpretations of the trial results, considering evidence from related studies	8 (7, 10)
58. Implications of trial findings to future research, policy, and practice	9 (7, 10)
<b>Other Information</b>	
59. Registration number and name of trial registry	9 (7, 10)
60. Ethical approval (if needed), informed consent procedures, and important ethical considerations	10 (8, 10)
61. References to all other sources of information about the methods and outcomes of this trial (full trial protocol, other papers or reports about the trial)	9 (8, 10)
64. The role of the intervention developer in the design, conduct, analysis, and reporting of the trial	9 (8, 10)
65. Any other potential conflicts of interest, including how they were managed	10 (8, 10)

### **3.3 Round 2 Results**

#### *3.3.1 Quantitative Rankings*

In Round 2, 321 participants recommend that 22 of the round's 66 items should be included in the checklist; a list of all 58 items from both rounds that were recommended for inclusion in the CONSORT-SPI checklist is provided in Table 7. The remaining 44 items were ranked indeterminate (Table 8). Based on *a priori* criteria for consensus, no items were excluded using Delphi Round 2 rankings. A checklist of items after Round 2 mapped onto the potential CONSORT-SPI checklist can be found in Appendix O.

#### *3.3.2 Text Responses for Items Recommended for Inclusion in Round 1*

On the whole, most participants made no comments on or strong objections to items that were recommended for inclusion in the CONSORT-SPI checklist during Round 1: “Very interesting to see the consistency on these items - nothing further to recommend on these” (P056).

Some concerns were raised about the potential size of the checklist already: “36 seems a lot of items! I assume some will be merged at the face-to-face meeting?” (P015).

The impossibility of blinding participants and providers in this area was reiterated, with suggestions about what the guidance should recommend under these items: “For a behavioral intervention in which the participant and provider cannot be blinded to the intervention, you want the investigators to report how potential bias was handled, PARTICULARLY (capitals for emphasis not shouting) if a self-reported outcome measure is used” (P173).

Another theme was that the recommended process evaluation items were important but not always relevant to all trials in this area, and even when relevant may be quite lengthy to discuss in a paper: “I still find the process evaluation section tricky...this is

something I always include in detail in my cluster RCTs of public health interventions but writing it up really requires a separate paper to do it justice. Also I'm aware that many trials are just not able to do this and not only eg for budgetary reasons" (P203).

Some commented that they did not think sample size calculation is needed as a checklist item, with concern about statistical items already in CONSORT: "I propose that all reportees using p-values (or any random sampling derivatives) must state clearly what the probabilities are probabilities of. They must also persuade the reader that their cases are truly random (i.e. complete and with no dropout/refusal). Finally, they must clarify on each occasion whether their concern is with generalisation to the population of cases originally allocated to groups or whether it is of that population to other cases that took no part in the study" (P017).

A final theme was aiming to identify unique features for social and psychological intervention RCTs throughout the Delphi process and CONSORT-SPI checklist as a whole: "these issues are not specific to SPIs, just important to any trial. What seems to be key is identifying those items that are specifically important/more important for trials of SPIs" (P114).

### *3.3.3 'Title and Abstract' Text Responses*

The proposed items on the title and abstract in Round 2 were considered helpful for scanning articles, particularly as a systematic reviewer: "Both title and summary should be able to allow the readers decide if the paper is important for a entire appreciation, once there are so many papers and people neither allways [sic] have time to examine all articles they have access" (P052). However, there were still concerns about overly-prescriptive recommendations for the title: "I am concerned that titles will become unwieldy with too many wording requirements. It is more important to indicate that the study was

randomized; allow for other important information to be found in the abstract” (P383).

There was also a desire to keep abstract items to a minimum: “These items overburden the abstract, especially in instances where a trial has many arms and outcomes” (P013).

### 3.3.4 “Introduction” Text Responses

For items on the introduction, “framing” comments were provided to orient the project team as to what is important to recommend in this section of the manuscript: “General negative reaction is that inclusion of these items will bloat papers in areas in which there has been significant research and will make authors strain to include other more critical items and stay within page limitations issued by various journals” (P289). Moreover, any new items in the introduction section could just be extensions of the “objective/hypothesis” item in CONSORT: “Justification for each objective and hypothesis - I think that this information should be implicit from “the scientific background and explanation of rationale of the study” (P152).

Concern was raised about using the term “evidence-based” in the checklist to describe interventions: “Definitions of “evidence-based interventions” vary by the criteria used by alternative rating organizations” (P025). Rather, participants stressed the need to cite previous systematic evidence on an intervention in the introduction in order to justify the trial: “These reporting guidelines should support the field's consolidated attention on this issue, by demanding greater accountability of researchers in building systematically on the existing evidence base” (P167).

There were issues again with the wording of the item on how the intervention is supposed to work, as a full conceptual framework may be “too much” to report for some interventions: “On conceptual framework, I think something is needed; but logic model/conceptual framework are fairly specific things. Perhaps could be made more

general” (P066). Suggestions were to focus more on mechanisms in this item: “I think what we want is the mechanism of action for the intervention and to ensure that the mechanism(s) are tested as part of the outcome measures” (P171).

There was concern that giving clusters a special status throughout the checklist may not be appropriate for a minimum, user-friendly set of reporting items: “if considering a minimum set, specifically indicating whether objectives related to the individual or to a cluster need not be necessary (although data analysis section should make clear about cluster analysis if appropriate)” (P280).

### *3.3.5 “Trial Design, Randomisation, and Blinding Procedures” Text Responses*

Issues were again raised about items related to blinding procedures: “I think it is a given that certain forms of blinding which are viable in drug trials are not for complex interventions, so wouldn' [sic] want to have to make this point every time a study was written up” (P224). Some had issues with the term “blinding” itself, instead preferring to use another term like “masking”: “Used in this context, the terms "blind" and "blinding" are objectionable to some (many) people. Perhaps use the terms "mask" and "masking"? I am reluctant to support any item that uses the term "blind" or its extensions” (P025).

It was also indicated that describing control conditions and their differences with the intervention may be covered by the CONSORT checklist and other items in the Delphi rather than the item proposed for this specific section: “Last item (similarities and differences) may be covered by items describing the detail of the intervention(s) and comparator(s)” (P306).

### *3.3.6 “Trial Setting” Text Responses*

There was support for increased attention to setting in the CONSORT-SPI checklist:

“I think you need all of the above for people to be able to determine how generalisable the results are, and what resources they would need to implement the intervention” (P152). Still, there were requests to change the phrasing of items in this section: “The problem with most of these items is the weasel word 'relevant'; the assessment of the relevance of contextual influences is likely to be personal to the author(s)” (P166). There were also issues with reliable measurement and reporting of certain aspects of setting: “For many community based and large scale area based interventions it is more likely that, without conducting further policy and local research, what will be available will be fairly rudimentary ... superficial impressions” (P172). As such, some felt that this area of research may not be at a sufficient stage to determine the critical aspects of context: “I’m unsure about these because what would be relevant seems so variable across topics. They would have to be stated in ways that make sense across a big range” (P038). Some concrete suggestions were provided for information on timings of key study procedures, though: “Timings should be in the flow diagram. Dates should be included in the text- allows for the calculation of recruitment rates, and a better methodological understanding for researchers using similar methods in similar populations” (P078).

### 3.3.7 *“Interventions” Text Responses*

Much of the information requested by proposed items was considered to be potentially more relevant for secondary papers: “Most of this content points to a important point. Not all information could be offered in one publication. For many of these aspects, a reference to a comprehensive study protocoll [sic] might be enough. However, this opportunity should also be mentioned in the CONSORT-SPI statement” (P075).

There were also voiced concerns about items related to mediation testing as well as knowledge about contextual influences: “Causal mechanisms will depend on the theory

being used. From my perspective, the intervention is designed to examine the cause of change because it is the only thing being manipulated. Any other mechanism (e.g., measured variable examined for mediation) is speculation by association” (P033).

Another theme was concern that items are not as applicable to “higher-level” interventions: “Some of these items do not consider how or whether the questions could be applied to large scale social policy interventions, for example transport interventions, welfare interventions, housing improvement etc. I therefore do not consider that these items should be considered as mandatory and will only serve to irritate authors and potentially prevent publication [sic] of very valuable research” (P172).

### 3.3.8 “*Outcomes and Data Analysis*” Text Responses

For methods related to outcomes and data analysis, some Round 1 items that participants requested to be split items in Round 2 were now being requested to be combined for the checklist: “Many items here could be merged, e.g, modify the first as follows: / Describe measures used (with references) in all reported analyses, including mode of administration, piloting, and any modifications to pre-existing measures” (P015).

Still, it was thought that greater clarity was needed about the goal of items on outcome data and analysis: “The issue here is about selecting the most valid and reliable tool. What we really want is to demonstrate that authors are / / a) clear about the outcome domain under investigation (e.g., depression), / b) clear that the selection of the outcome [sic] domain matches the treatment. Is there content in the treatment that is focussed [sic] on changing depression. / c) they select a measurement tool that is valid: i.e., able to capture and describe depression / d) that the measurement tool is sensitive to treatment change and insensitive to time (reliable) / e) that the measurement tool selected as the primary measure remains the primary measure” (P018).

There was support for an item on imputation, but discrepant suggestions on what to include in it: “Details on imputation are sadly lacking in many papers currently and this standard needs to be improved in light of new capacities in statistical packages so that the analytic descriptions are stronger” (P176).

There were also various views on the level of detail needed on psychometric properties of measures: “Psychometrics of well established measures should be referenced but not described. Psychometrics for any adaptation or of new measures should be provided or if reported in another publication referenced” (P101).

### 3.3.9 “Baseline and Outcomes Data” Text Responses

There were various views on phrasing the item on baseline characteristics: “Be specific about exactly what demographics should be reported: age, race, ethnic/cultural status, socioeconomic status, gender, AND presence of any existing psychological or mental health problems” (P198). However, despite high quantitative rankings, there was considerable cynicism in text comments about comparing study completers and all those providing data at baseline: “It is pointless comparing study completers with non-completers. They will be different so what? The important thing is to compare the groups as analysed to see if attrition altered the balance achieved at randomization” (P230).

There was also debate on which additional analyses need to be reported in the main trial paper: “Subgroup analysis often happens later and this requirement [sic] will delay reporting of main results” (P281). Though a potential solution to problems with choosing which additional analyses to report in a main trial paper, there was still concern that the field is not yet ready for raw data to be required from every trial: “Not sure about practical feasibility of providing raw datasets for all studies... as it is at the research team's discretion about how available they want their data to be and there may be some issues

relating to research funders etc which would make this requirement tricky” (P020).

As in Round 1, there was support for the reporting of adverse events, but scepticism as to whether an extended item is needed in this checklist: “Agree that adverse is important, but agree with comment that this is part of the 'harm' item in the CONSORT already” (P185). In addition, some suggestions were provided on how to identify and handle *a priori* analyses and methods versus those done post hoc: “Would prefer one paragraph ... outlining where the protocol differs” (P120).

### 3.3.10 “Process Evaluation” Text Responses

For the items related to the results of process evaluations, participants had opposing views about which provider characteristics are actually important to report: “I think training and supervision more important than credentials” (P197). There were even continued opinions that process evaluation results constitute a separate paper altogether: “I think these things are very important but should be addressed in a separate paper on process evaluation. They should be required reporting, but not necessarily in a single main effects paper” (P086).

There was also continued concern about analyses on mechanisms and contextual influences: “Causal mechanisms and contextual influences are complex enough topics that they will typically require separate papers. They are also matters of considerable methodological debate -- many methodologists would say that common methods are not acceptable” (P038). Participants also had some disagreement on the reliability of identifying barriers and facilitators of implementation: “Barriers are potentially useful, but often amount to educated guesses and unsure if needs to be included in CONSORT. Perhaps better to suggest how they should be handled to limit rampant non-empirical guessing” (P185).

### 3.3.11 *“Discussion” Text Responses*

For the discussion section, there was advocacy continued from Round 1 for less prescriptive standards: “I don't see a need for requirements about the discussion, if the requirements call for all the necessary information in earlier sections. If we require specifying the population, procedures, setting, control condition, etc. in the methods, I don't see a need for requiring it in the discussion” (P038). There was also a desire to target only what is possible to reliably say about limitations and generalisability: “By what criteria will applicability be judged? Perhaps, the reporting guideline could ask the authors of trials to specify what criteria they use to judge applicability to different populations and groups transparently” (P129). In addition, there was some debate about what implications should be discussed, if any: “Am reluctant to say that single studies should report policy implications. Could this be over-reaching the data? It seems to me that policy implications accrue over the course of many studies where the pattern of findings begins to have program implications” (P025).

### 3.3.12 *“Other Information” Text Responses*

Regarding other information, there was a minority of views against the proposed ethical approval item, particularly as it was worded: “link to ethics/IRB number should be sufficient along with ethical considerations, putting in the informed consent is overkill” (P144). There was, however, a fairly strong desire in text comments for there to be guidance, at least in the E&E document, about conflicts of interest: “Reporting of conflicts [sic] of interest seldom occurs. I have not known many researchers in social science areas, as opposed to medical /health trials, to actually think about this at all. Maybe not essential [sic] but worth drawing attention to this for social science interventions so that more consideration is given to this aspect” (P176).

There were also continued concerns that the field may not be ready for an item on trial registration: “There are many fields in which there is no trial registry. Unless there is one, the topic is irrelevant” (P038).

### *3.3.13 Text Responses for New Items Suggested in Round 1*

Regarding the new items proposed from Round 1, there were various views about what is important to report regarding contamination: “Perhaps the measurement of whether contamination actually occurred is even more motivated? Attempts to avoid or minimize contamination is most likely addressed by the study design and covered in the Methods section” (P247).

Another theme was that acknowledgements should be up to the authors rather than an item in the guideline: “Acknowledgements of all contributors could result in a very substantial list. Best to leave this up to authors discretion” (P101).

There was also a view that costs are important but should not be required in main paper: “Cost information should not be a requirement of main trial write-up. Obviously it is a very useful extra piece of information, but usually the economic evaluation will be published separately” (P020).

A specific item on power analyses may not be needed due to other items in the CONSORT 2010 checklist and proposed in the Delphi process: “Power [sic] analysis should not be required if the sample size estimation has been included earlier in the Method” (P033).

In contrast to the previous items, there were strong views in favour of an item on stakeholder involvement: “As the field bumps into massive challenges in moving from efficacy trials to the field - often without the benefit of effectiveness trials - it is critical that stakeholders are involved in trial design, conduct, and/or analyses and...later down the

road...in dissemination strategies. Hence, I vote to include this information” (P167).

### *3.3.14 Superordinate Themes from Delphi Process Text Responses*

Across the entire Delphi process, participants’ comments largely highlighted the desire for a minimal, user-friendly checklist that only includes standards applicable to all trials in this area. As such, the themes that came through the qualitative feedback quite strongly and consistently were: worries about the length and size of the checklist, to keep it a user-friendly document; requests to make the wording as helpful as possible; desire to pick and phrase items that are minimal and apply to *all* social and psychological intervention trials (e.g., individual and place-based interventions); and requests to be sensitive to the “phase” of the trial when giving advice about items (i.e., pilot vs. efficacy vs. effectiveness).

## **4. Discussion**

### ***4.1 Overall findings***

This Delphi process provided helpful and rich information for the CONSORT-SPI consensus meeting and guideline document write-up. Participants rated many items as highly important for social and psychological intervention trials. That is, most items had a strong consensus for inclusion in the CONSORT-SPI guideline documents, as indicated by high-rankings, low-dispersion of scores, and lack of significant free-text comments suggesting needed changes. The importance of many of these items is further substantiated by existing empirical research on the concepts underlying items, as well as existing reporting guidelines in the medical, behavioural, and social sciences that address these concepts.

**Table 7. Recommended Items for the CONSORT-SPI Checklist from the Delphi Process**

Section and Topic	Item	Median (IQR) from Round 1 or % from Round 2
<b><i>Title and abstract</i></b>		
	Identify as randomised/randomized in the title	10 (8, 10)
	The abstract should identify the population, all intervention and control conditions, outcomes of interest, times of follow-up, and the trial setting	I = 83%, E = 6%, U = 10 %
<b><i>Introduction</i></b>		
	Scientific background and explanation of rationale of the study	10 (8, 10)
	Describe the problem(s) or issue(s) that the intervention(s) is intended to address	10 (9, 10)
	Mention current knowledge about the effectiveness of the experimental intervention (e.g., reference previous systematic reviews)	I = 90%, E = 5%, U = 4%
	Specific objectives or hypotheses of the study	10 (9, 10)
<b><i>Methods</i></b>		
<b><i>Trial Design</i></b>		
	Description of trial design (such as cluster, factorial, crossover), including allocation ratio	10 (9, 10)
	Report all inclusion and exclusion criteria for participants, providers, settings, and (if relevant) clusters	10 (9, 10)
	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	10 (9, 10)
	How sample size was determined	9 (8, 10)
	When applicable, explanation of any interim analyses and stopping guidelines	8 (7, 10)
	Incentives offered to participants (e.g., to enrol in the trial, use the intervention, complete outcome measures)	I = 84%, E = 4%, U = 12%
	Rationale for choice of the control/comparator intervention(s) in the trial	I = 85%, E = 9%, U = 6%
<b><i>Randomisation</i></b>		
	Method used to generate the random allocation sequence	9 (7, 10)
	Type of randomisation (e.g., minimisation, stepped-wedge) and details of any restriction (such as blocking and block size)	10 (8, 10)
	Mechanism used to implement the random allocation sequence (such as sequentially numbered opaque envelopes), describing any steps taken to conceal the sequence until interventions were assigned	9 (7, 10)

<b>Section and Topic</b>	<b>Item</b>	<b>Median (IQR) or %</b>
<i>Randomisation (cont.)</i>	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	9 (7, 10)
<i>Blinding Procedures</i>	Whether and how providers and participants were blind after assignment to interventions, and if maintenance of blinding was assessed	9 (8, 10)
	Whether and how outcome assessors were blind after assignment to interventions, and if maintenance of blinding was assessed	10 (8, 10)
	Description of similarities and differences in content and delivery between intervention(s) and comparator(s)	I = 86%, E = 5%, U = 9%
<i>Setting</i>	Dates/timings of study procedures by trial arm (e.g., recruitment, baseline, intervention, and follow-up)	I = 82%, E = 8%, U = 10%
	Geographic location of the trial (e.g., rural setting in Southwest US, urban setting in London, UK)	I = 89%, E = 2%, U = 8%
<i>Interventions</i>	Precise details of the content of the intervention(s) and comparator(s) as designed for the study, including clear definitions of the essential and non-essential components for all groups, and the intended differences across groups	10 (9, 10)
	Precise details of the intended duration and frequency of the intervention(s) and comparator(s)	10 (8, 10)
	Precise details of the intended format of the intervention(s) and comparator(s), such as individual vs. group, in-person vs. electronic provision	10 (8, 10)
	Describe how and by whom actual delivery and uptake of the intervention by providers and participants was assessed	I = 81%, E = 7%, U = 12%
<i>Outcomes and Data Analysis</i>	Explain the choice of outcomes, their timing and length of follow-up, and any differences across groups in how outcomes are measured	9 (8, 10)
	Any changes to trial outcomes after the trial commenced, with reasons	10 (8, 10)
	Statistical methods used to compare groups for primary and secondary outcomes, with reasons	10 (9, 10)
	Methods for additional analyses, such as subgroup analyses, adjusted analyses, and how these compare to the trial registration and protocol	9 (8, 10)
	Imputation methods for handling missing data, and whether these methods were pre-specified	9 (8, 10)
	Describe measures used in all reported analyses, including mode of administration and any modifications to preexisting measures	I = 91%, E = 2%, U = 7%

<b>Section and Topic</b>	<b>Item</b>	<b>Median (IQR) or %</b>
<i>Outcomes and Data Analysis (cont.)</i>	Identify all outcome measures in the trial, and if these match the trial registration/protocol	I = 86%, E = 6%, U = 8%
	Any transformations or changes to raw quantitative data, with reasons	I = 86%, E = 5%, U = 9%
<b><i>Results</i></b>		
<i>Participant Flow and Recruitment</i>	A flowchart including the following for each group: the numbers of participants, clusters, and providers or centres who were (1) approached, (2) screened, (3) eligible, (4) randomly assigned, (5) received the intended intervention, and (6) were analysed for the primary outcome, including the number of participants by each provider or center and reasons for dropout	10 (9, 10)
	For each group, losses and exclusions after randomisation, including the number of participants who discontinued the intervention but remained in the trial, together with reasons	10 (9, 10)
	Whether the trial has ended or was stopped, with reasons if so	10 (8, 10)
	For each group, number of participants (denominator) included in each analysis, and whether each analysis was per protocol or based on initial intervention assignment	10 (8, 10)
	Data for demographic, socioeconomic, and other participant characteristics measured at baseline per trial arm	I = 93%, E = 1%, U = 6%
	Describe whether and how the study completers differ from the original sample on baseline characteristics per trial arm	I = 87%, E = 8%, U = 5%
<i>Process Evaluation</i>	Precise details of the intervention(s) and comparator(s) actually offered by providers, with reasons for any differences from design	9 (8, 10)
	Precise details of any tailoring by providers of the intervention(s) and comparator(s) to individual participants across groups	9 (7, 10)
	Precise details of the intervention(s) and comparator(s) actually taken up by participants	9 (7, 10)
	Amount of the intervention(s) and comparator(s) actually received by participants (e.g., sessions attended) across groups	9 (8, 10)
<i>Outcomes and Estimation</i>	For each quantitative outcome, the results for each group as well as the estimated effect size and its precision (such as 95% confidence interval)	10 (9, 10)
	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	9 (8, 10)

<b>Section and Topic</b>	<b>Item</b>	<b>Median (IQR) or %</b>
<i>Outcomes and Estimation (cont.)</i>	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	10 (8, 10)
<b><i>Discussion</i></b>		
	Interpretation of the results, considering pre-specified and alternative hypotheses	I = 93%, E = 4%, U = 3%
	Population(s) to whom the results may apply, considering sample characteristics, the intended population, recruitment procedures, and related studies	I = 82%, E = 9%, U = 10%
	Limitations due to sources of potential bias and imprecision	I = 95%, E = 2%, U = 2%
	Limitations due to heterogeneity (e.g., variability in the participants, intervention implementation, and outcomes)	I = 84%, E = 8%, U = 7%
	Limitations of methods to investigate intervention implementation (e.g., provider delivery and participant uptake)	I = 82%, E = 12%, U = 7%
	Implications of trial findings to future research, policy, and practice, commensurate with strengths and limitations of the study	I = 85%, E = 7%, U = 8%
<b><i>Other Information</i></b>		
	References to intervention manual(s), websites, and other resources concerning the intervention	9 (7, 10)
	Sources of funding and other support, and the role of funders in the design, conduct, analysis, and reporting of the trial	10 (8, 10)
	References to external sources with information about the methods and outcomes of this trial, such as the full trial protocol and other papers/reports about the trial	I = 82%, E = 9%, U = 9%
	Report any involvement of the intervention developer in the design, conduct, analysis, and reporting of the trial	I = 81%, E = 8%, U = 10%
	Any potential conflicts of interest	I = 95%, E = 1%, U = 3%

I = Include, E = Exclude, U = Uncertain

**Table 8. Indeterminate Items from the Delphi Process**

<b>Percent Include</b>	<b>Section and Topic</b>	<b>Item</b>	<b>Round 2 Percentages</b>
<b>75-79%</b>	<i>Other Information</i>	Details about ethical approval (body giving approval/study identification number, informed consent procedures, any important ethical considerations)	I = 79%, E = 9%, U = 12%
	<i>Methods: Outcomes and Data Analysis</i>	Methods used to enhance the quality of measurements in the trial (e.g., multiple observations, training of assessors)	I = 79%, E = 8%, U = 13%
	<i>Methods: Outcomes and Data Analysis</i>	If done, what variables were used for imputation	I = 79%, E = 7%, U = 12%
	<i>Results: Outcomes and Estimation</i>	Distinguish pre-specified from exploratory analyses	I = 79%, E = 8%, U = 12%
	<i>Other Information</i>	How any identified potential conflicts of interest were handled	I = 78%, E = 8%, U = 14%
	<i>Results: Outcomes and Estimation</i>	Any adverse psychological events or unanticipated social disadvantage to individuals or clusters (e.g., families, communities)	I = 77%, E = 9%, U = 13%
	<i>Discussion</i>	Interpretation of any results about moderators and mediators, considering other relevant evidence and study limitations	I = 77%, E = 12%, U = 12%
	<i>Discussion</i>	Setting(s) to which the results may apply, considering intervention implementation, choice of comparator, trial context, and related studies	I = 77%, E = 11%, U = 12%
	<i>Other Information</i>	Registration number and name of trial registry, or reasons why trial was not registered	I = 77%, E = 10%, U = 13%
	<i>Methods: Interventions</i>	Any methods to avoid or minimise contamination or spillover effects in the trial (e.g., participants in a comparator group receiving the experimental intervention)	I = 76%, E = 10%, U = 13%
<i>Methods: Outcomes and Data Analysis</i>	If done, the number of imputations performed	I = 75%, E = 8%, U = 16%	
<b>66-74%</b>	<i>Title and Abstract</i>	The intervention and the target problem/population should be identified in the title	I = 72%, E = 15%, U = 12%
	<i>Results: Outcomes and Estimation</i>	Report or refer readers to documents reporting analyses additional to main effects (e.g., subgroups; adjusted analyses)	I = 72%, E = 12%, U = 16%
	<i>Methods: Blinding Procedures</i>	Discuss reasons for lack of blinding at any stage (e.g., impossible to blind participants to their assigned intervention)	I = 71%, E = 16%, U = 13%
	<i>Methods: Blinding Procedures</i>	Whether and how data analysts were blind after assignment to identifying information about participants	I = 70%, E = 16%, U = 14%
	<i>Methods: Interventions</i>	Describe the plan for staff training and support for the intervention(s) and comparator(s)	I = 69%, E = 16%, U = 15%

<b>Include</b>	<b>Section and Topic</b>	<b>Item</b>	<b>Round 2 Percentages</b>
<b>66-74%</b>	<i>Introduction</i>	Provide a conceptual framework or logic model for how the intervention is hypothesised to lead to changes in outcomes	I = 68%, E = 13%, U = 20%
	<i>Methods: Interventions</i>	Any methods to investigate intervention causal mechanisms (qualitative/quantitative exploration of hypothesised mediators)	I = 68%, E = 13%, U = 19%
	<i>Methods: Interventions</i>	Describe any piloting of the intervention(s) with providers and the implementing organisation(s) prior to commencing the trial	I = 68%, E = 13%, U = 18%
	<i>Results: Process Evaluation</i>	Actual training to deliver the intervention(s) (rather than per the study protocol)	I = 67%, E = 16%, U = 17%
	<i>Methods: Blinding Procedures</i>	Any methods to address lack of blinding (e.g., participants unaware content of comparator to minimise demand characteristics)	I = 67%, E = 15%, U = 19%
	<i>Results: Process Evaluation</i>	Details of any other identified barriers and facilitators of implementing the intervention(s)	I = 67%, E = 14%, U = 19%
	<i>Other Information</i>	Acknowledgements of those who contributed substantially to the project but did not meet authorship requirements	I = 66%, E = 17%, U = 16%
<b>50-65%</b>	<i>Methods: Trial Design</i>	Stakeholder involvement in trial design, conduct, and/or analyses (e.g., practitioners, policymakers, participant representatives)	I = 65%, E = 15%, U = 20%
	<i>Methods: Outcomes and Data Analysis</i>	Reference(s) to any validation studies for each measure, noting comparability of their population(s) and context(s) to the trial's	I = 65%, E = 13%, U = 21%
	<i>Results: Process Evaluation</i>	Results of any investigations of causal mechanisms of the intervention	I = 64%, E = 16%, U = 20%
	<i>Methods: Outcomes and Data Analysis</i>	Information about each measure's psychometric properties in the trial (e.g., internal consistency of self-report measures)	I = 64%, E = 15%, U = 21%
	<i>Introduction</i>	Identify any evidence-based interventions for this problem/issue and how the experimental intervention differs	I = 63%, E = 19%, U = 18%
	<i>Results: Process Evaluation</i>	Results of any investigations of contextual influences on intervention outcomes	I = 62%, E = 17%, U = 22%
	<i>Introduction</i>	Justification for each objective or hypothesis	I = 61%, E = 18%, U = 21%
	<i>Results: Process Evaluation</i>	Actual supervision of providers across groups (rather than per the study protocol)	I = 59%, E = 20%, U = 21%
	<i>Methods: Interventions</i>	Any methods to investigate contextual influences on intervention outcomes (e.g., qualitative/quantitative exploration of the dependence of outcomes on practice setting, implementing organisation, or external environment)	I = 59%, E = 17%, U = 23%

Include	Section and Topic	Item	Round 2 Percentages
50-65%	<i>Introduction</i>	Whether each objective or hypothesis pertains to the individual participant level and/or cluster level (e.g., family, community)	I = 55%, E = 21%, U = 24%
	<i>Methods: Setting</i>	Characteristics of practice setting(s) directly related to experiences of participants and providers (e.g., provider/participant ratio, physical space to run intervention)	I = 55%, E = 18%, U = 26%
	<i>Results: Outcomes and Estimation</i>	Present results of power analyses (i.e., power for outcome analyses at each follow-up)	I = 52%, E = 27%, U = 20%
	<i>Results: Process Evaluation</i>	Actual professional qualifications of providers (rather than per the study protocol)	I = 52%, E = 24%, U = 24%
	<i>Results: Process Evaluation</i>	Describe any information about the acceptability or perceived value of the intervention(s) and comparator(s) by participants	I = 52%, E = 21%, U = 27%
	<i>Methods: Interventions</i>	Describe the planned physical and technical resources for the intervention(s) and comparator(s)	I = 51%, E = 23%, U = 25%
	<i>Methods: Interventions</i>	Describe the plan for staff recruitment and selection for the intervention(s) and comparator(s)	I = 50%, E = 27%, U = 22%
	<i>Methods: Setting</i>	Characteristics of the larger implementing organisation(s) that shape the practice setting (e.g., private/public school ownership, competing priorities to prison-based intervention)	I = 50%, E = 19%, U = 31%
<50%	<i>Methods: Blinding Procedures</i>	Whether maintenance of blinding data analysts was assessed	I = 49%, E = 26%, U = 25%
	<i>Methods: Setting</i>	Characteristics of the external environment relevant to the trial (e.g., community demographics, health/social care policies)	I = 45%, E = 23%, U = 32%
	<i>Results: Outcomes and Estimation</i>	Provide raw data set needed for replicating analyses (either in online supplement or indicate how to obtain data on request)	I = 44%, E = 30%, U = 26%
	<i>Methods: Interventions</i>	Direct costs of implementing the intervention (e.g., provider salary, necessary material resources, training and supervision)	I = 29%, E = 40%, U = 31%

I = Include, E = Exclude, U = Uncertain

It is valuable to note that all but one of the items from the CONSORT 2010 Statement reached consensus for inclusion in the CONSORT-SPI guidelines during Round 1. Overall, this demonstrates that the CONSORT Statement is a useful tool for trials in the social and behavioural sciences in addition to the biomedical sciences, and that building off the extensive previous work of the CONSORT Group is a viable solution for improving the quality of social and psychological intervention trials. The only CONSORT checklist item to not receive consensus in Round 1 was the need to provide a trial registration number; free-text comments revealed this is likely due to a perception that trialists in this area by and large do not register trials in advance, which has been verified in empirical studies.<sup>1</sup>

Many items currently not in CONSORT guidance have been recommended for inclusion in the checklist, such as those to do with theoretical mechanisms of the intervention and intervention implementation. Some items that were indeterminate at the end of the Delphi process did receive some support when these items were relevant to a trial (e.g., certain aspects of process evaluation), though these items were consequently not deemed necessary to report in every trial. Still, such results indicate that the reporting quality review from the previous phase of the project led to the selection of high quality, pertinent items to consider for the CONSORT-SPI checklist, as no items were generally seen as unimportant in the Delphi process.

#### ***4.2 Strengths and Limitations of the Current Study***

There are several strengths and limitations of this study, as is often the case with Delphi processes given their flexible and consequently highly variable implementation. Firstly, according to an on-going systematic review led by the current author, this is the largest Delphi process ever conducted for developing a reporting guideline. The achieved

sample size of this Delphi process is a major strength due to previous issues of buy-in and take-up of CONSORT guidance by behavioural and social scientists. This large size was also needed due to the heterogeneity of disciplines targeted by the CONSORT-SPI guideline, and thus the need to have these disciplines well-represented. Such heterogeneity of stakeholders may not be the case with other guidelines targeting a more focused population. Interestingly, this heterogeneous group did largely hold the same views on items, as indicated by the low dispersion of most (if not all) scores. Furthermore, the quota for journal editors was reached, as well over 25% of participants were journal editors. In addition, the high retention rate of over 80% in Round 2 indicates good engagement with these targeted stakeholders. The breadth of disciplines, countries, and career stages indicate a potential variety of perspectives that will inform the development of the final guideline documents.

One potential limitation of this study is the high proportion of participants with a background in psychology, and from those who reside in the US or the UK. However, both psychological researchers as well as US and UK research networks do produce a disproportionately large amount of RCTs in this area, giving a rationale to sampling such a large number of stakeholders from these backgrounds.

On a similar note, there was a low recruitment rate at the approach and screening stage of this study, possibly due to the number of journal editors approached that may not publish social and psychological intervention RCTs. This yields the possibility that those entering the Delphi process were predominantly researchers in support of CONSORT or reporting guidelines more generally, and thus are likely to share similar views that may differ from the wider population of trialists in this area. Furthermore, at times some participants reported that they didn't understand terms or items. Without a glossary, some participants noted, in the feedback during the process, that some technical terms (e.g.,

fidelity) may have been interpreted in an idiosyncratic fashion, complicating interpretation of results.

Another thing to note is that no items were actually excluded during this Delphi process. Like the Delphi process for the SPIRIT Statement, this Delphi process yielded a large number of items that were recommended for inclusion in the checklist during considerations at a subsequent consensus meeting.<sup>12</sup> Also akin to the SPIRIT Statement Delphi process, such a large amount of items could be partly due to the significant number of items provided in the Round 1 questionnaire, which had already been reviewed by the project's IAG. Items not recommended for inclusion were likely seen as important or ideal to report when relevant, but may not be appropriate for a minimum checklist: for example, because they do not apply to all social and psychological intervention RCTs. This state of affairs left the consensus meeting participants—who were actually charged with deciding the content of the CONSORT-SPI checklist—to deliberate more and make tougher decisions about guideline content than if items had been excluded. To address this issue, many items recommended via the Delphi process may need to be subsumed under the items in the ultimate CONSORT-SPI checklist, as well as discussed in greater detail in explanatory papers where appropriate.

The ability to consider the relative importance of items to each other was potentially limited by the sequential, emergent presentation of items in the questionnaire format. As such, participants had to have done a second pass through the whole questionnaire to adjust ratings according to items' relative importance. As many participants likely did not make a second pass, this format may have led to the highly skewed nature of the questionnaire.

#### ***4.3 Implications and Future Directions***

This Delphi process generated a prioritised list of items to consider at the

CONSORT-SPI consensus meeting. The participants at the consensus meeting used this list, along with results from the systematic reviews in the previous thesis chapter, to extend the CONSORT 2010 checklist to trials of social and psychological interventions.

Qualitative feedback from the Delphi process was also particularly useful for developing the Explanation and Elaboration document for the CONSORT-SPI Extension Statement.

The rationing needed to develop a minimal checklist from these responses, and the opportunities to discuss reporting standards in the E&E, were the goals of the consensus meeting discussed in the next chapter.

## References

1. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.
2. Grant S, Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo-Wilson E. Developing a reporting guideline for social and psychological intervention trials. *Research on Social Work Practice*. 2013;23(6):595-602.
3. Montgomery P, Grant S, Hopewell S, et al. Protocol for CONSORT-SPI: An Extension for Social and Psychological Interventions. *Implementation Science*. 2013;8:99.
4. Mayo-Wilson E, Grant S, Hopewell S, Macdonald G, Moher D, Montgomery P. Developing a reporting guideline for social and psychological intervention trials. *Trials*. 2013;14:242.
5. Gardner F, Mayo-Wilson E, Montgomery P, et al. Editorial Perspective: The need for new guidelines to improve the reporting of trials in child and adolescent mental health. *Journal of Child Psychology and Psychiatry*. 2013;54(7):810–812.
6. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
7. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998;2(3).
8. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*. 2000;32(4):1008-1015.
9. Boukdedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review.

- PLoS One*. 2011;6(6):e20476.
10. McKenna HP. The Delphi technique: A worthwhile research approach for nursing? *Journal of Advanced Nursing*. 1994;19:1221-1225.
  11. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *The Lancet*. 2008;371:281-283.
  12. Tetzlaff JM, Moher D, Chan AW. Developing a guideline for clinical trial protocol content: Delphi consensus survey. *Trials*. 2012;13.
  13. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BJOG: An International Journal of Obstetrics and Gynaecology*. 2013;120(6):765-770.
  14. Keeney S, Hasson F, McKenna H. *The Delphi technique in nursing and health research*. Oxford: Wiley-Blackwell; 2011.
  15. Montgomery P, Grant S, Hopewell S, Macdonald G, Moher D, Mayo-Wilson E. Developing a reporting guideline for social and psychological intervention trials. *British Journal of Social Work*. 2013;43(5):1024-1038.
  16. Grant S, Montgomery P, Hopewell S, Macdonald G, Moher D, Mayo- Wilson E. Letter to the Editor: New guidelines are needed to improve the reporting of trials in addiction sciences. *Addiction*. 2013;108:1687-1688.
  17. Grant S, Mayo-Wilson E, Hopewell S, Macdonald G, Moher D, Montgomery P. Developing a reporting guideline for social and psychological intervention trials. *Journal of Experimental Criminology*. 2013;9(3):355-367.
  18. Mayo-Wilson E, Montgomery P, Hopewell S, Macdonald G, Moher D, Grant S. Developing a reporting guideline for social and psychological intervention trials. *British Journal of Psychiatry*. 2013;203:250-254.
  19. Montgomery P, Mayo-Wilson E, Hopewell S, G M, Moher D, Grant S. Developing

- a reporting guideline for social and psychological intervention trials. *American Journal of Public Health*. 2013;103(10):1741-1746.
20. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311(376).
  21. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*. 2011;8(1):e1000393.
  22. Montgomery P, Grant SP. Development of a CONSORT extension for complex social interventions. Paper presented at: Meeting at the 20th annual Cochrane Colloquium. 2012, October; Auckland, New Zealand.
  23. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
  24. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: Explanation and Elaboration. *PLoS Medicine*. 2008;5(1):e20.
  25. Fitch K, Bernstein SJ, Aguilar MD, et al. *The RAND/UCLA Appropriateness Method User's Manual*. Santa Monica, CA: RAND; 2001.
  26. Taylor WJ, Schumacher Jr HR, Baraf HS, Chapman P, Stamp L, Doherty M. A modified Delphi exercise to determine the extent of consensus with OMERACT outcome domains for studies of acute and chronic gout. *Annals of the Rheumatic Diseases*. 2008;67:888-891.
  27. Froud R, Eldridge S, Kovacs F, et al. Reporting outcomes of back pain trials: A

- modified Delphi study. *European Journal of Pain*. 2011;15(10):1068-1074.
28. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Huberman AM, Miles MB, eds. *The qualitative researcher's companion*. Thousand Oaks, CA: Sage Publications; 2002:305-330.
  29. Bowling A. *Research methods in health: Investigating health and health services*. New York: McGraw-Hill; 2009.
  30. Corbin J, Strauss AC. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London, UK: Sage Publications; 2009.
  31. *NVivo qualitative data analysis software* [computer program]. Version 10. 2012.
  32. Liamputtong P. Qualitative data analysis: conceptual and practical considerations. *Health Promotion Journal of Australia*. 2009;20(2):133-139.
  33. Malterud K. Qualitative research: Standards, challenges, and guidelines. *Lancet*. 2001;358(9280):384-388.
  34. Mays N, Pope C. Qualitative research in health care: Assessing quality in qualitative research. *BMJ Clinical Research*. 2000;320(7226):50-52.

**Chapter 5:**  
**The CONSORT-SPI Consensus Meeting:**  
**Developing the Guideline Checklist**

**Abstract**

**Background:** Following a systematic review of the literature and a Delphi process, experts in the development of research reporting guidelines recommend holding a face-to-face meeting of key stakeholders in the area targeted by the proposed guideline, in order to determine the content of the guideline checklist. The purpose of this chapter is to report on the consensus meeting discussions related to the selection of items for the CONSORT-SPI checklist.

**Methods:** A group of 31 researchers, journal editors, and funders met in March 2014 to extend the CONSORT 2010 Statement to RCTs of social and psychological interventions. A three-day consensus development conference was held to discuss preliminary research on social and psychological intervention RCTs, vote on items for the CONSORT-SPI reporting standards checklist, and discuss a dissemination and implementation plan for the CONSORT-SPI guidelines.

**Results:** The group recommended that the CONSORT-SPI checklist consist of 14 extended items of the CONSORT 2010 checklist, as well as a flow diagram for tracking participants through an RCT. Participants at the meeting also provided several suggestions for defining social and psychological interventions, rewording and repositioning extended checklist items, and possibly renaming the guideline to better engage those researchers working on interventions that target systems or environments.

**Conclusions:** The CONSORT-SPI checklist should facilitate better reporting of social and psychological intervention trials. Transparent reporting of the planning of and discussions during the CONSORT-SPI consensus meeting should also provide valuable insights for future reporting guideline development teams.

## 1. Introduction

A formal consensus process is a crucial aspect of developing a high-quality reporting guideline. The Delphi method serves as a helpful consultation of stakeholders for prioritising items for consideration in a reporting guideline checklist.<sup>1</sup> Following a Delphi process, experts in the development of research reporting guidelines recommend holding a face-to-face meeting of key stakeholders in the area targeted by the proposed guideline, in order to actually determine the content of the guideline checklist. A face-to-face meeting is suggested in order to help elucidate reasons for held opinions (which are harder to clarify without in-person interaction), and to consider the most important or contentious reporting items in greater depth.<sup>2</sup>

In previous reporting guideline projects, a two- or three-day formal consensus meeting is usually held to determine guideline *content*, rather than wording or format.<sup>3</sup> A few days are typically allotted to allow sufficient time for thorough discussion, reducing stressful and hasty decision-making that can hinder judgment.<sup>2</sup> Results from the systematic review and Delphi process should inform consensus meeting discussion, as participants debate what should be included in the guideline and how to disseminate the resultant documents. Therefore, a consensus development process was used for the actual selection of checklist items for the CONSORT-SPI extension so that participants could confirm the results of the Delphi process and discuss the opinions raised in greater detail. However, to date, the actual planning for—as well as detailed minutes of—these consensus meetings to develop reporting guidelines have not been thoroughly described in guideline documents.<sup>4</sup> Lack of transparency about consensus meeting methodology makes it difficult for readers to appraise the decisions made about guideline content, and also makes it difficult for future reporting guideline developers to learn from previous reporting guideline projects.

The purpose of this chapter is to summarise the discussions at the consensus meeting

that led to the selection of items in the CONSORT-SPI checklist. This thesis chapter describes these discussions in detail and presents the resultant draft of the CONSORT-SPI checklist decided by meeting participants. This report will serve as an initial foundation for the CONSORT-SPI checklist document.

## **2. Methods**

### ***2.1 Eligibility Criteria***

Participants were drawn from the Delphi process and selected purposively to include stakeholders from various disciplines and various professional roles, namely trialists, funders, and journal editors.<sup>4</sup> At the outset, a group of around 20-30 participants was desired to balance the diversity of opinion with meaningful opportunities for interaction,<sup>5</sup> and to be sufficiently small to maximise opportunities for discussion and create an opportunity to achieve consensus.<sup>2</sup>

### ***2.2 Recruitment Strategy***

The Project Executive first identified a small group of invitees from the IAG, whose participation was deemed essential to hold the meeting.<sup>5</sup> After this small group was invited, and potential dates for the meeting had been settled, members of the IAG were again consulted to finalise the size of the consensus meeting group and the participants to invite. Based on recommendations from the IAG and other members of the Project Executive, invitees were sent formal invitation emails, tailored to their current level of involvement in the project. For example, members of the IAG themselves were given a brief email asking them to attend, whereas those invitees who had (at most) only participated in the Delphi process received a longer email with published project manuscripts as attachments. If a member of the IAG or Project Executive knew the invitee well, this IAG or Project Executive

member also sent the invitee a note of encouragement to attend. A list of people larger than the target number was kept in case invitees declined, or those who had accepted had to withdraw from the meeting. In such instances, someone from the list with the closest disciplinary background and professional role(s) was then approached. Invitations were sent from one to six months in advance of the meeting. Travel, accommodation, and meals for each participant were provided or reimbursed.<sup>5</sup>

### ***2.3 Time and Setting***

The meeting was held from 17 to 19 March 2014 at the Hawkwell House Hotel in Oxford, UK. Meeting discussions took place in one large room, and all discussions were as one group; three tables were arranged around two projector screens, on which presentations were displayed. The vast majority of participants stayed on site, barring a few who lived in Oxford.

### ***2.4 Meeting Preparation***

The Project Executive developed the meeting agenda in advance, allowing an adequate form and structure to the meeting in order to facilitate the proper amount of time needed for all of the agenda items.<sup>5</sup> The primary goal of developing an early version of the CONSORT-SPI checklist guided the development of the agenda.

Pre-meeting reading packs were also developed and sent to participants approximately three weeks in advance of the meeting. These materials included short professional biographies of all participants, the agenda for the consensus meeting (Appendix L), and a brief summary of the CONSORT-SPI Delphi process. They also included five published manuscripts related to the project: guidance for the development of reporting guidelines,<sup>5</sup> the conceptual launch paper for the CONSORT-SPI project,<sup>6</sup> the review of reporting quality and

reporting guidelines related to social and psychological interventions,<sup>7</sup> the CONSORT-SPI project protocol,<sup>8</sup> and the CONSORT 2010 Statement E&E to give participants an example of the end goal of this meeting.<sup>9</sup> The appendices included a draft CONSORT-SPI checklist based on the Delphi results (Appendix P), final rankings for items from the Delphi process, and a summary of the Delphi Round 2 qualitative feedback. The same materials were also provided during the meeting (in a printed pack and on a flash drive), with the addition of the newly published Template for Intervention Description and Replication (TIDieR) Checklist and manuscript,<sup>10</sup> as well as the CONSORT Extension for Abstracts.<sup>11</sup>

Pre-meeting phone calls were conducted by one member of the Projective Executive with each consensus meeting attendee. These pre-meeting calls aimed to make sure that all logistical issues had been sorted, see if participants had any questions about the agenda and reading materials, discuss with each participant their expectations of the meeting and the discussions therein, and to identify additional issues to add to the agenda.

### ***2.5 Structure of Meeting***

The consensus meeting followed established methods<sup>3</sup> used in previous CONSORT meetings.<sup>12-15</sup> Members of the CONSORT Group, as well as several people who have participated in previous consensus meetings, were also in attendance to facilitate an efficient meeting. As mentioned above, literature reviews and results of the Delphi process were provided in advance to ground conversations on empirical information and to facilitate cohesive discussion.<sup>2</sup> The meeting itself involved presentations on relevant background literature, sharing the results of Delphi process, discussion of each proposed item, voting following item discussion, and sessions focusing on the strategy for producing documents and disseminating project outputs. PowerPoint slides of these presentations will be made available on the project website.

At the beginning of each day, the objectives for that day were reviewed, and any outstanding issues among the participants were addressed. Throughout all three days, the meeting was divided into sessions, with members of the Project Executive rotating as discussants of each session to lead discussion and to ensure that decisions were made when applicable to a session.<sup>5</sup> While the Project Executive aimed to follow closely with the agenda, the actual length and timings of sessions were somewhat flexible so as to adapt to the needs of the group. An Administrative Assistant was also hired to assist with the computer screens, logistics at the site, and other assorted tasks for smooth running of the meeting.

### *2.5.1 Day 1 Agenda*

The consensus meeting began with an introduction of all participants, followed by an overview of the goals of the meeting, and then several presentations about specific background topics related to reporting social and psychological intervention trials.<sup>12</sup> These presentations included a discussion on what reporting guidelines are, systematic reviews on reporting quality and reporting guidelines related to social and psychological intervention RCTs, the Project Executive's conceptualisation of social and psychological interventions, and the results of the CONSORT-SPI Delphi process.

Two structured sessions solely involving group discussion were also on the agenda: one after the presentation on the systematic reviews to have participants discuss reporting problems in their area, and another after the presentation on the Delphi process to have participants discuss its results. Background presentations and discussion on Day 1 were considered a useful way to have participants become acquainted with each other and to set the stage for the discussions and voting on items during Days 2 and 3.<sup>5</sup> By the conclusion of this day, all participants were intended to understand the current quality of reporting social and psychological intervention RCTs, the general results of the Delphi process, and the scope

of the guideline.

### *2.5.2 Day 2 Agenda*

At the beginning of the day, participants were reminded in an overview session about Day 2 that the group was voting on items for the CONSORT-SPI checklist, and should consider the items included in the checklist to be a minimum, essential set reflecting standards that should be reported based on empirical evidence of their importance to bias, previous poor reporting, and/or theory.<sup>5</sup> Classifications as to the importance of selected items are noted in the analysis of consensus meeting discussions below.

Participants were led in a structured discussion of each item proposed for the checklist from the Delphi process.<sup>12</sup> The list of items at the end of the Delphi process was revised for the consensus meeting based on both quantitative rankings and qualitative responses from Delphi participants. Consensus meeting participants discussed whether each proposed checklist item should be included or modified for the final version of the CONSORT-SPI checklist. As much as possible, these discussions focused on the concept underlying each item rather than seeking to perfect wording at this stage, although differences in nomenclature were noted so as to inform the guideline documents.

For each session, the discussant facilitated and ensured that all participants were able to express any views, that all ideas were discussed in-depth, and that assertive participants did not dominate the discussion.<sup>2</sup> Participants then voted on the minimum items to include in the CONSORT-SPI checklist. Voting was confidential using anonymous electronic ballots to promote honest answers and allow participants to rethink their position if a re-vote was needed.<sup>16</sup> Proposed items appeared on one screen, and instructions to vote on the other screen (i.e., press “A” for Include, “B” for Exclude”, or “C” for unsure as to whether an item should be included in the CONSORT-SPI checklist). Items were voted on individually in turn, and

participants were notified that they had the ability to change their votes if they pressed the wrong button. After the meetings for Day 2 were completed, the Project Executive compiled a draft CONSORT-SPI checklist for discussion on Day 3.

### *2.5.3 Day 3 Agenda*

After an overview of the objectives of Day 3 and a session on tabled issues from Day 2, there was a lengthy session reviewing and discussing the items recommended for the CONSORT-SPI checklist from the previous day. The overview of items selected from the previous day also involved discussion on the content of the CONSORT-SPI flow diagram, using the CONSORT for Non-Pharmacologic Treatments flow diagram as a template.<sup>17</sup> The Project Executive originally planned to have separate sessions on reviewing and discussing recommended items (bit by bit) and then finalising the CONSORT-SPI checklist as a whole (with a global view), but these were decided on the morning to be collapsed into one session because the group felt that both could be done at the same time. A new session on the appropriateness of the guideline title, “social and psychological”, then replaced the session on finalising the CONSORT-SPI checklist because of requests from some participants. Once the checklist items had been decided, participants discussed and then voted on proposed items for the title and abstract. A subsequent 40-minute session on discussion of the checklist wording was planned, but this was cut because the group’s conversation had already addressed checklist wording. Remaining issues in wording will be handled by the Project Executive and writing group before circulating a draft checklist to the consensus group.

The meeting concluded with discussion about optimising knowledge translation, and members of the group committed to specific efforts to this end.<sup>12</sup> Namely, the strategy for producing guideline documents, as well as the strategy for disseminating and implementing these documents, were discussed.<sup>5</sup> Though originally planned to be two sessions, the

discussion on implementation from the viewpoint of editors and the publication strategy session were merged into one session. This combination provided more time for the session on knowledge transfer. The discussions specifically about dissemination and implementation of CONSORT-SPI guidance will be in Chapter 7 of the thesis. The focus of this chapter is to report the discussions leading to the first draft of the CONSORT-SPI checklist and E&E.<sup>18</sup>

## ***2.6 Data Analysis***

When voting on items for the first time (Days 1 and 2), any items with  $\geq 70\%$  “include” votes from participants were considered to be included in the CONSORT-SPI checklist. Any item with 70% “include” and “unsure” rankings combined were allowed to be raised again on Day 3 at the Project Executive’s discretion. Anything with  $\geq 30\%$  “exclude” would not be raised again on Day 3, unless strongly desired by one or more consensus meeting participants. On any Day 3 re-votes, items with  $\geq 80\%$  “include” rankings were included in the CONSORT-SPI checklist, anything with  $\geq 70\%$  “include” could be added at the discretion of the writing group at a later date, and anything with  $< 70\%$  “include” was considered to have insufficient support to be considered a minimum reporting item. The required high rates of consensus aimed to ensure that the checklist consisted of a minimum set of essential items for reporting social and psychological intervention trials.<sup>2</sup>

All meeting discussions were audio-recorded. Interpretation of consensus meeting discussions will be presented in a narrative form, with the approach taken limiting the narrative to content of dialogue as it progressed from session to session.<sup>19</sup> Narratives are not meant to be an exact record or mirror of an experience;<sup>19</sup> rather, the focus of this analysis is an interpretative summary encapsulating the main discussions underlying the decisions made during the consensus meeting. Rather than fracturing such accounts into decontextualised pieces, narrative presentation allows an organised understanding of the sequential, structural

features of dialogue, making such an approach particularly useful when participants organise thoughts and replies as extended accounts or conversations.<sup>20</sup> As such, the interaction amongst participants will be noted as and when it helps to elucidate the author's interpretation of statements made by participants. It is worth noting that the DPhil candidate perceived the meeting to involve friendly, mutually-respectful discourse, as interactions amongst participants were usually to express agreement, elaborate on similar views, or to voice disagreement in a professional manner and often acknowledging different disciplinary perspectives.

The analysis first began with attending to the meeting dialogue, which involved listening to and transcribing verbatim the entirety of the content of meeting discussions from the audio-recordings of each session. After the audio was transcribed, the DPhil candidate thoroughly read and re-read the material in order to obtain an overview of the conversations at the meeting beyond impressions gained while attending the meeting. Then, to structure the interpretation of the consensus meeting conversations, the transcript was coded line-by-line according to categories that were grounded in the data itself,<sup>21</sup> using NVivo version 10 to facilitate the process.<sup>22</sup> Paragraphs or several paragraphs at a time were coded to preserve the full content of responses, any pertinent interaction amongst participants,<sup>23</sup> and minority or alternative views on particular subject matters.<sup>2</sup> After this first round of coding, notes taken by the DPhil candidate during the meeting were used to re-code the transcripts, focusing the narrative on those aspects of the conversation salient to understanding the remit of the CONSORT-SPI guideline, decisions about its content, and suggestions for its implementation and dissemination.<sup>19</sup>

Following this stage, a loose oral narrative summarising the discussions was constructed, creating a "metastory" signifying what occurred during the meeting via an edited reshaping of participant discussion.<sup>24</sup> The narrative is organised below into sessions of the

consensus meeting. Participants have been given anonymised identifiers, with quotations selected that illustrate the content and tone of discussions. While coding was only performed on the content of the discussions rather than providing line notes about tone and context, interpretations were based on the audio recording, as well as the DPhil candidate's impressions and notes from the meeting, to increase the credibility of the analysis. The narrative is restricted to those comments pertinent to summarising participants' views on the CONSORT-SPI checklist, E&E document, and dissemination; discussions related to informal comments as they emerged were not analysed and are not included.

### **3. Results**

#### ***3.1 Flow of Participants and Items***

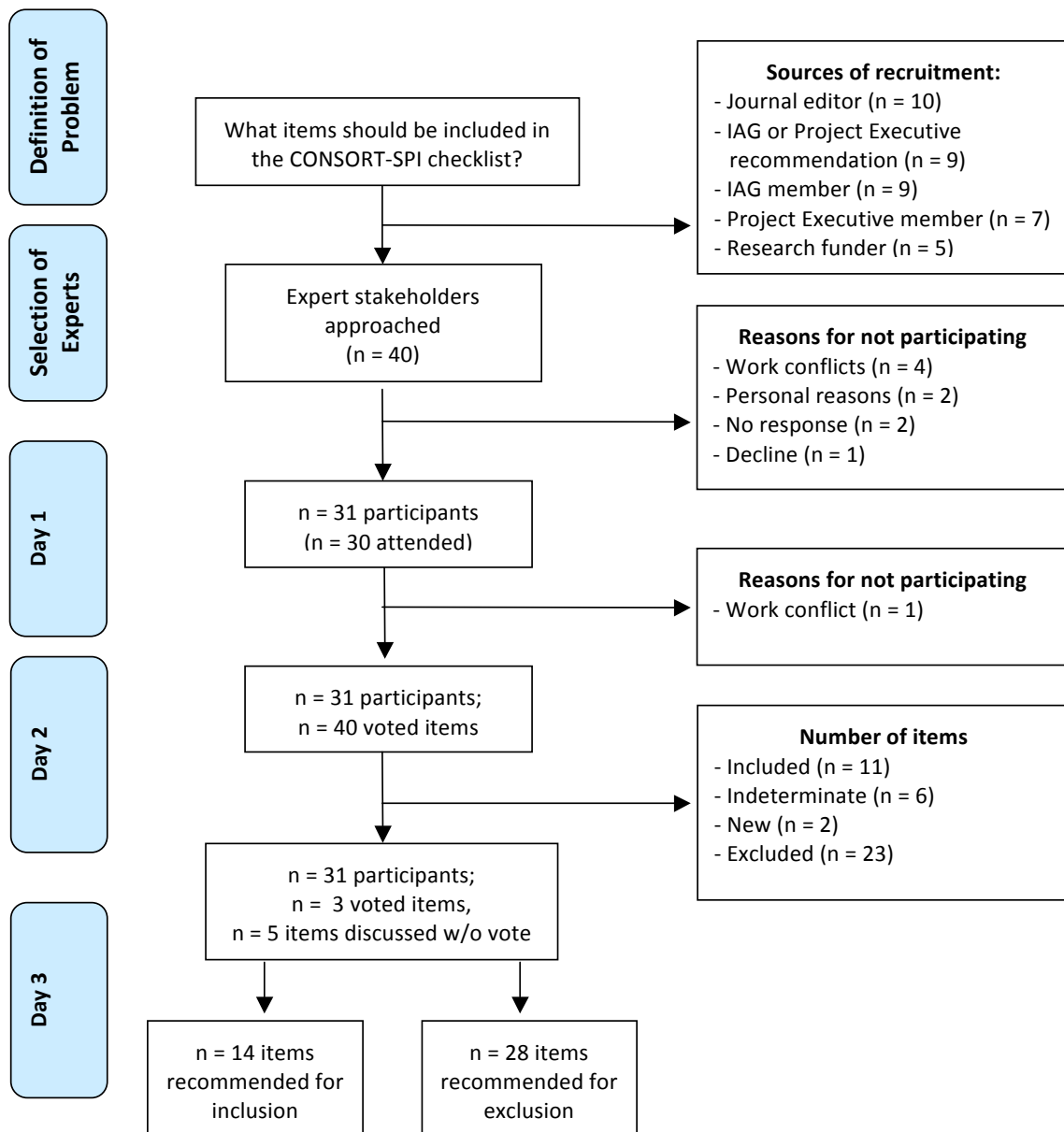
The flow of participants and items through the study can be found in Figure 3. Overall, 31 participants attended the consensus meeting, including the seven members of the Project Executive, eight members of the IAG, nine journal editors, three representatives from research funding organisations, and four intervention research methodologists. There were 42 items considered during the consensus meeting: 40 items were discussed on Day 2, and then 6 of these (in addition to two items for the title and abstract) were discussed on Day 3. Overall, 14 items were recommended for inclusion in the CONSORT-SPI Checklist, and 28 items were not recommended for inclusion (for rankings for each item, see Table 1). A full checklist of the CONSORT-SPI items at the end of Day 2 and at the end of the consensus meeting can be found in Appendices Q and R, respectively.

#### ***3.2 Day 1 Meeting Minutes***

##### ***3.2.1 Session 1.1—Welcome & Introductions (Discussant P5)***

This first day began with a welcome and introduction of every person in attendance at

**Figure 1. CONSORT-SPI Consensus Meeting Participant Flow Diagram**



**Table 1. Votes on Proposed CONSORT-SPI Checklist Items during the CONSORT-SPI Consensus Meeting**

CONSORT 2010 Item #	Proposed CONSORT-SPI Checklist Item	Percent Include	Percent Exclude	Percent Unsure	Percent No Response	Percent Error
2a	The problem(s) or issue(s) targeted by the intervention(s)	60.0	36.7	0	3.3	0
2a	Current knowledge about the intervention	40.0	56.7	3.3	0	0
2b	Whether or how the intervention is hypothesised to work	73.3	23.3	3.3	0	0
3a	If the unit of random assignment is not the individual, please refer to CONSORT for Cluster Randomised Trials	93.3	6.7	0	0	0
None	Stakeholder involvement in trial design, conduct, and/or analyses (e.g., practitioners, policy-makers, participant representatives)	73.3	16.7	6.7	3.3	0
None	When applicable, how intervention providers were assigned to each group	90.0	3.3	3.3	3.3	0
11a	Awareness of intervention assignment by providers, participants, and assessors throughout the trial	56.7	43.3	0	0	0
11b	Similarities as delivered and received	23.3	76.7	0	0	0
5	TIDIER framework labels in the checklist	60.0	40.0	0	0	0
5	Incorporate TIDIER points into the E&E (Brief name, why, what, who provided, how, where, when, how much and how well, tailoring, modifications, planned versus actual, etc.)	96.7	3.3	0	0	0
4b	Important features of intervention setting(s), delivering organisation(s), and wider context	60.0	36.7	3.3	0	0
5	How and by whom actual delivery and uptake of the intervention was assessed	63.3	33.3	3.3	0	0
5	Extent to which intervention and control conditions were delivered and taken up as planned	96.7	3.3	0	0	0
6a	Choice of outcome measures and length of follow-up AND Mode of administration, modifications to preexisting measures, and differences across groups outcome measurement	16.1	83.9	0	0	0

CONSORT 2010 Item #	Proposed CONSORT-SPI Checklist Item	Percent Include	Percent Exclude	Percent Unsure	Percent No Response	Percent Error
6b	Identify all outcome measures in the trial, and if these match the trial registration/protocol	19.4	80.6	0	0	0
12a	Rationale for analytic methods	29.0	64.5	3.2	3.2	0
12a	Imputation methods for handling missing data and any transformations or changes to data, with reasons	67.7	29.0	0	3.2	0
12b	Whether analyses match the trial registration/protocol	61.3	32.3	3.2	3.2	0
4a	Eligibility for providers and settings as applicable	45.2	51.6	0	3.2	0
4a	Rationale for choice of the control/comparator	51.6	45.2	0	3.2	0
4a	<b>Incentives offered as part of the trial</b>	90.3	6.5	0	3.2	0
13a	<b>Where possible, the number approached, screened, and eligible prior to random assignment, with reasons for dropout (and adjust flow diagram accordingly)</b>	80.6	16.1	0	3.2	0
13b	Number of participants who discontinued the intervention but remained in the trial	19.4	77.4	0	3.2	0
14a	Dates and timings of study procedures by trial arm	64.5	29.0	3.2	3.2	0
14b	Indicate if trial is still ongoing at time of publication	22.6	74.2	3.2	0	0
15	<b>Include socioeconomic variables where applicable</b>	74.2	22.6	0	3.2	0
15	Differences between those who provide outcome data at follow-up and original sample per trial arm	32.3	64.5	0	3.2	0
17a	Power to detect postulated effect for each analysis of primary outcomes	32.3	67.7	0	0	0
18	Indicate other results from trial, with references if available	54.8	41.9	0	3.2	0
18	<b>Indicate availability of trial data</b>	77.4	19.4	0	3.2	0
19	Adverse psychological events or unanticipated social disadvantage to individuals or groups	6.5	90.3	0	3.2	0
20	Limitations related to heterogeneity and intervention implementation	19.4	74.2	0	6.5	0
21	Population(s) and setting(s) which the results may apply and those which they may not	6.5	83.9	0	6.5	3.2

CONSORT 2010 Item #	Proposed CONSORT-SPI Checklist Item	Percent Include	Percent Exclude	Percent Unsure	Percent No Response	Percent Error
22	Implications for future research	22.6	71.0	0	6.5	0
23	If not registered, reasons why (OR: Indicate if registration is a priori)	19.4	74.2	0	6.5	0
24	Where the intervention manual and other informational materials can be accessed	90.3	6.5	0	3.2	0
24	Where other information about the trial's methods and outcomes can be accessed	45.2	51.6	0	3.2	0
25	Any involvement of the intervention developer in the design, conduct, analysis, and reporting of the trial	80.6	12.9	3.2	3.2	0
25	Allegiance to the intervention	22.6	71.0	0	6.5	0
25	Any potential conflicts of interest and how they were handled	19.4	74.2	0	6.5	0
*12a.	Statistical approaches for missing data are explicitly stated	83.9	3.2	0	12.9	0
*TBD	Declaration of any other potential interests	80.6	3.2	3.2	12.9	0
*1a	Intervention and population/problem in title	51.6	22.5	19.4	6.5	0

The above table is organised chronologically by the order in which participants voted on items.

“CONSORT 2010 Item #” refers to the item number of the CONSORT 2010 checklist to which each CONSORT-SPI checklist item was proposed to correspond at the time of the vote. A “\*” next to an item number denotes a vote that took place on Day 3; the rest took place on day two.

“Proposed CONSORT-SPI Checklist Items” are worded in the table above as they were worded on the screen during the vote.

**Green items** have currently been voted for **inclusion** in the draft CONSORT-SPI checklist.

**Red items** have currently been voted for **exclusion** in the draft CONSORT-SPI checklist.

the consensus meeting, as well as members of the project's International Advisory Group who were not attending the meeting. These introductions set to establish a respectful, receptive group atmosphere for discussion. Following this introduction, a member of the Project Executive ran through some house-keeping items, such as the plans for dinner that evening, as well as asking participants to sign the consent form for the meeting and audio recording had they not completed it yet. After checking whether anyone had questions about those house-keeping items (which none did), the session leader then decided to move on to the session about goals and processes.

### *3.2.2 Session 1.2—Overview of Goals and Process (Discussant P13)*

The discussant began with an overview of the processes leading up to the meeting, the purpose of the consensus meeting, and the next stages in the CONSORT-SPI project after the consensus meeting. The discussant explained that the purpose of the consensus meeting was to think about any issues specific to social and psychological interventions that might not be covered by the CONSORT 2010 Statement.<sup>14</sup> The discussant also informed the participants that the Project Executive hoped this group would help with the dissemination of the guideline once completed and that, as the discussions continued, participants would think about specific ways that they could do so.

The discussant then reviewed the five phases of the CONSORT-SPI project. The discussant informed the group that the presentations later on Day 1 would review the completed phases (i.e., Phases 1 and 2), emphasising that the checklist items to be considered at the consensus meeting came from these phases: “It is that process that has led to the list of items that we have in front of us today. So those haven't come out of nowhere; those are items that were identified during the lit reviews and then now have been looked at by a very large group of people and led us to the meeting today” (P13). The

discussant then indicated that the project was currently at “Phase 3”—deciding draft CONSORT-SPI checklist items at the consensus meeting. The group was informed Day 1 was meant to provide a thorough understanding of the preliminary work from Phase 1 and 2, and then Days 2 and 3 were meant to identify a draft list of items for the CONSORT-SPI checklist.

On Day 3, the group would also spend time trying to think about the publication strategy as well as optimal dissemination methods. The discussant informed participants in particular about the desire to have members of the group assist with developing an Explanation and Elaboration (E&E) Document<sup>9</sup> for each of the various disciplines represented at the meeting. These E&E documents would indicate how items related to social and psychological interventions in each targeted area, and would provide examples of good reporting specific to each of these areas as well. The group was also informed of the potential for spin-off, “educational” publications about methodological issues in conducting RCTs of social and psychological interventions: “Coming out of the Delphi process, ... [a] number of people ... don’t yet understand how to register a trial in each of the different disciplines that are represented here.... That’s one of the areas that ... we’re going to need to work with you on to maybe write some editorials or to do some presentations at conferences to increase understanding” (P13). The discussant suggested such publications could also arise after the consensus meeting for items deemed important but currently may not have sufficient evidence to include in the checklist, so that an evidence base could be developed in time for a potential update to CONSORT-SPI. It was emphasised that the publication and dissemination process would not be “overnight”, but rather something only beginning post-meeting.

The discussant then informed the group that, as for the Delphi process, participants at the consensus meeting were purposively selected as trialists, intervention researchers,

research funders, practice guideline developers, policy-makers, and practitioners across key disciplines of relevance to the project. The goal was to have as many pertinent perspectives represented, given the intended widespread impact of the guideline, so that participants could voice the needs of relevant communities to which they belong: “we see this guideline as impacting people in all of these different areas and we wanted to make sure that the recommendations that we have and the guidance that we have really meet the needs of all of those different groups” (P13).

To conclude this session, the discussant informed the group of the project outputs to date, namely a launch paper (and shorter editorials) co-published in several journals that invite stakeholders to participate in an inclusive guideline development process. Those journal editors in attendance were asked to think about their interest in considering any of this project’s outputs for (co-)publication once complete. The discussant then briefly mentioned previous conference presentations about CONSORT-SPI, expressing the desire for those interested to think about presenting on CONSORT-SPI at relevant conferences in their area down the line.

### *3.2.3 Session 1.3—Background: Reporting Guidelines (Discussant P17)*

After an initial greeting and welcome, the discussant began by describing how the reporting guidelines movement arose in the early 1990s, partly in response to a research article on the lack of reporting “negative” trials. The early movement culminated in the 1996 version of the CONSORT Statement. The discussant then went on to define a reporting guideline to provide a shared understanding of the term for the rest of the meeting: “I think essentially, it’s a checklist. It’s a checklist that could be considered to some degree as a memory jog for what to include in a report of a randomised trial” (P17). The discussant went on to emphasise that this is a *minimum* checklist rather than an all-

inclusive list. While items not in a checklist may still be important (e.g., ethical concerns are not included in the CONSORT 2010 checklist<sup>14</sup>), checklist items should be particularly unique to the methodology targeted by a guideline. Items should also have some evidence to support the importance of their inclusion (e.g., not reporting information is associated with bias). The discussant further emphasised the concept of a minimum checklist in light of the numerous items under consideration from the CONSORT-SPI Delphi process; the discussant suggested that the group aim to have a checklist that fits on one page so as to avoid barriers to its subsequent adoption and implementation.

The “audience” of a reporting guideline was then discussed. Primarily targeted to authors, the discussant mentioned other important stakeholders such as peer-reviewers, editors, and research funders. Broadening the group’s understanding of stakeholders transitioned into broadening the group’s understanding of the “CONSORT Family”, proposing to the group that they should be mindful of how CONSORT-SPI will exist in conjunction with existing CONSORT guidelines, as well as other reporting guidelines. One new guideline in particular, the Template for Intervention Description and Reporting (TIDieR),<sup>10</sup> was then briefly highlighted, given that several CONSORT-SPI consensus meeting participants were involved in developing it, in addition to its relevance to many items in the CONSORT-SPI Delphi process on reporting interventions.

The discussant then turned to the EQUATOR Network’s recommendations for developing reporting guidelines,<sup>5,25</sup> a process that CONSORT-SPI was following. The discussant underscored the importance of thinking about dissemination and implementation throughout the entire development process. This includes having all guideline documents and outputs being open access to lower barriers for authors and to facilitate use, evaluating the impact of the guideline, and getting it properly endorsed by journals. Participants were advised to consult the EQUATOR Network’s library of

reporting guidelines to see examples of guideline development and dissemination, as well as to understand the wider context of reporting guidelines within which CONSORT-SPI will exist. Before opening the floor to discussion, participants were shown a slide from a recent review on the impact of CONSORT,<sup>26</sup> demonstrating the positive impact a reporting guideline can have as an “intervention” on reporting quality, yet also revealing that there is still room for much improvement.

One participant asked the discussant to return to a slide on existing CONSORT extensions, questioning the coherence of how all of these extensions can be conceptually organised. Namely, the participant was the first at the meeting to propose the important task that the group had in defining social and psychological interventions for the purposes of CONSORT-SPI:

I was thinking about what does “social and psychological” interventions mean? And it might be helpful to see if we have a shared understanding of that. It could be where the dependent variable, the target is a psychological mechanism or a social process. It could be interventions where one thinks the mechanisms of change are social or psychological, or it could be that the intervention itself is social or psychological. So my question is: is it all three of those? Or one? Or two of those? How are we defining it?  
(P32)

Another member of the group responded in part to this question, leading to a brief interaction about the term non-pharmacological to describe interventions:

P1: The big box to the left of the three on the right, is the non-pharmacological interventions, which is actually highly pertinent, though of course not the same thing. It would include surgery, for example. But I can’t answer your last question either.

P32: OK, just if I can throw in another comment, a particular bug-bear of mine is using the word “non-pharmacological”. It’s like talking about psychological/social: let’s call them non-medical.

P1: Oh I totally agree. I hate defining things by what they are not, but it’s extremely hard to think of a word that covers everything except drugs.

P32: Or to think of a word that covers everything apart from behaviour. And maybe that’s not what we need to do.

The discussion next focused on the proposed length of the checklist. One participant warned that a forensic, detailed approach for items would lead to a checklist with “dozens

of pages” (P1). Another responded that, while understanding that barriers to implementation may exist when guidelines are seen as bureaucratic box-ticking exercises, a more detailed form to fill in while writing a paper could be extremely helpful:

On the one hand you don't want to be too prescriptive .... but I think it's worth considering certainly, having some level of detail, and quite an extensive set of guidance, on many of those items that people could then use, and constructing it in a way that people, in a Word document, that almost it writes the paper for them. (P15)

The discussant replied that such detailed information could go into the E&E paper, and that, given the amount of evidence on poor reporting of research, being prescriptive to a certain extent is well-justified. Another participant suggested that items be distinguished as primary or secondary for the time being, so as to avoid an artificial limit of a one-page checklist until the group began to determine what items truly were important in this area (P29). A participant who is both a journal editor and graduate professor responded in favour of considering a longer checklist due to concerns that any details placed in the E&E document will not be as thoroughly attended to as those items on the checklist itself. Having such details in the checklist itself would further suggest the importance of the details for trial conduct as well as reporting (P20).

One participant then raised the issue of prescriptiveness as it relates to maximising engagement and influence. A concern was raised about the reverence that many stakeholders hold towards social and psychological interventions, and how this must be considered when developing and ultimately disseminating the guidelines:

That there's something a bit mystical, or magical, or sophisticated about [social and psychological interventions] ... and it's almost as a threat to their identity to try to force them into this kind of checklist idea or into rigid guidelines. So I think we just need to bear in mind that context and the nature of one part of the audience that we're trying to get to when we present it. I think it's a question of framing and how we present it, and the language, talking about complexity, and really tackling the issue, and tackling in quite a sophisticated way, if we're going to achieve that engagement and take-up. (P32)

Another participant then acknowledged the salience of this insight, making the concluding point of this session that ultimately engagement will depend on buy-in by journal editors:

“if the journal editors get buy-in, people will shift what they do” (P26).

### *3.2.4 Session 1.4—Existing Guidelines and Reporting Quality of RCTs: Results from Systematic Reviews (Discussant P3)*

This session primarily involved a formal presentation on the two systematic reviews on existing reporting guidelines related to and the reporting quality of social and psychological intervention RCTs.<sup>7</sup> As these reviews are discussed elsewhere in the thesis (see Chapter 3), this narrative will only focus on points emphasised during the presentation, as well as any participant discussion.

At the outset of the presentation, participants were instructed to take out the physical hard copy of the paper<sup>7</sup> on these reviews as reference for the presentation. The discussant then reiterated that the CONSORT-SPI project was following the EQUATOR Network guidance for developing a reporting guideline, with these reviews as the first phase to make an empirical case of the need for CONSORT-SPI. The discussant highlighted that reporting standards from this review were the basis of items proposed in the first round of the subsequent Delphi process. During the discussion of reporting quality, the discussant also noted that some fields (e.g., clinical psychology) published more RCTs than others (e.g., criminology), and thus one potential avenue of work as part of this project might be to encourage more researchers to conduct trials in currently under-represented areas. The discussant also focused on a few specific topics when discussing reporting quality, highlighting the poor reporting of what was actually delivered and received by intervention *and comparator* participants, and then providing some examples of better and worse reporting amongst trials in the dataset. The importance of being clear about terminology was then addressed, as the discussant indicated the need ultimately to develop a glossary or taxonomy of terms for the guideline, while stressing that the purpose of the consensus

meeting was to decide checklist item *content* rather than word-smithing.

The first question raised in this session related to whether details provided in sibling documents (e.g., a trial protocol) were considered as reported or not in the systematic review on reporting quality (P29). The discussant responded that these were not coded as reported, yet the number of papers that referenced a linked document were quite minimal anyway, so deciding how to link documents on the same trial might be an important area to discuss during the consensus meeting.

As the participant posing the question indicated the increasingly common practice of publishing trial protocols (P29), another participant recommended explicitly mentioning the SPIRIT Statement<sup>27</sup> for developing and reporting trial protocols: “SPIRIT is interesting ... [because] it moves us away from prescribing reporting; though reporting is really important, clear reporting ... is at the end of the process. We want to influence something much more upstream” (P17). A few participants (P1, P17) indicated that CONSORT and SPIRIT items overlap quite a bit, leading another to ask how this guideline and CONSORT in general link back to PRISMA<sup>28</sup> guidance for reporting systematic reviews (P25). One participant mentioned the “cross-pollination” of CONSORT and PRISMA group members (P31). Another suggested SPIRIT, CONSORT, and PRISMA exist in the “parallel universes” of primary and secondary studies, yet the participant posing the question probed this view in light of content targeted by CONSORT-SPI:

But in many ways they're not [in parallel universes] because sometimes users of reviews are really keen to understand implementation and context and all of that sort of thing, but if it's not there in the primary studies, it doesn't make it into the reviews. So it has to be really tight I think between all three of the statements. (P25)

This point prompted another participant to suggest that these guidelines need to be cross-referenced, for example via hyperlinks within text of open access documents, to create “portals” between these “parallel universes” (P30).

Returning to the review slides, one participant questioned the quality of one of the

examples of “good reporting” of context that was provided during the presentation. The participant was concerned that adherence to items (related to, for example, context and theory) in CONSORT-SPI should not reward “good story-tellers”, but rather should seek to reveal the risk of bias and to increase the ability to replicate the trial (P18). Another participant asked whether the study investigated journals’ peer-reviewing guidelines, given the importance of peer-reviewers to the objectives of the CONSORT-SPI project (P9). A participant who had just published a study on reviewer guidelines in medical journals indicated that most journals don’t have quality reviewer guidelines available online, and some journals that do have quality guidelines provided by their publishers don’t even use them (P1).

Regarding the wider area of reporting guidelines, one participant mentioned that existing guidelines in social and behavioural sciences are written at a much higher “level of abstraction”, targeting empirical research in general, and that CONSORT-SPI (once completed) should be referenced in these (P33). The Journal Article Reporting Standards (JARS) of the American Psychological Association (APA)<sup>29</sup> were offered as one such example (P20). One participant followed with a few unrelated comments: that reporting quality in medicine is improving but still about as poor as the data cited in the review, that CONSORT traditionally leans more towards internal rather than external validity but this guideline could focus more on external validity, and that the group should be mindful not to veer CONSORT-SPI guidance from reporting to focusing on conduct of trials (P1). Support was provided for focusing on external validity, for example, via Glasgow’s RE-AIM model<sup>30</sup> (P14). This same participant then raised concerns about addressing differential response to social and psychological interventions due to the “agency” of the people engaged in the intervention. These comments were followed by a request to develop a “Rosetta Stone” for the terms being discussed in order to provide a shared

language for people from different backgrounds (P32). This participant then built on the point about external validity, arguing that RE-AIM works at a general level, but the field is still at the starting blocks in terms of theorising the nitty-gritty of context. A member of the Project Executive clarified that wording would be addressed during the writing phase of the project, and that the consensus meeting group will be consulted to generate a shared language.

Another participant requested that the group avoid preconceived notions of the importance of “external validity”, finding it much more subjective than reports of trial methods (P6). The final remark of the session considered the importance of historical perspective in deciding checklist items:

My own experience ... is that a lot of people, myself included on some occasions, have felt that we probably, our studies would have not met the criteria as you would have needed, because we kind of felt it was obvious.... But I think it is important, because if you don't say ... whatever it might be, people don't know what you did, so they don't know that it's sort of obvious that you did it ok. And I think that the lesson there is for all of the things. (P15)

### *3.2.5 Session 1.5— Group Discussion: Reporting Problems in “Your” Area (Discussant P13)*

The discussant began with an overview of the aims for the session: to have people from the various disciplines represented indicate the key problems in their areas, and identify common problems across the wider area of social and psychological intervention research. Participants involved in criminology research began the conversation. One noted that criminology is increasingly involved in the study of place-based or event-based interventions, as well as field experiments, which have implications for reporting RCTs in this area (P9). This participant then highlighted a particularly influential study in their area with a host of issues, such as selective outcome reporting, and not flagging the introduction of new variables in results papers that were not in the protocol. The discussant proposed

that many of these problems may be familiar to others, though the place-based issue might deserve unique attention. To clarify what was meant by “place-based” to one of the participants (P26), the examples of hot-spots police stops and anti-poverty programmes targeting public housing units were provided (P9, P12).

The discussion then turned to social work. Detail about interventions was immediately flagged as an important issue, given the dependence of the intervention on the behaviours of those involved (P30). Descriptiveness of procedures in the control group was deemed equally important, as well as providing data in a format that can be used by a systematic reviewer to conduct meta-analyses (P16). One participant seconded the importance of describing comparators, particularly highlighting problems with use of the term “treatment as usual” to describe what was provided (P5). Another participant noted in general the difficulty of speaking about the “social work literature”, as many social workers published in journals that are categorised under other disciplines, such as psychology and education (P28). Regardless, this participant voiced concerns about distinguishing descriptive from explanatory research, linking conceptual models with measurement models in studies, poor measurement of and analyses related to context, and again issues of measuring intervention and control condition implementation. These points invoked a brief interchange about statistical analyses related to context:

P13: One of the things I think that reminds me, this issue about context is in some ways about internal validity and some ways about external validity. So if you’ve got school as a dummy variable, you might be able to control it in one analysis, but to what extent the contextual factors facilitate take-up and make points that P32 and P5 were raising is another issue.

P28: Yeah and every time you add a higher-order variable like that, it costs you a degree of freedom, so you need to be very, very precise in your thinking about what are the contextual influences that you think are going to operate on an intervention, but we’re not at that place.

Researchers in education then provided their views. Previous issues were deemed important in education, particularly the description of the counterfactual condition as

“treatment as usual” (P33). Participant attrition and description of sampling procedures were also flagged as common problems in education—particularly at various levels of design in cluster trials—as well as the problem of “over-alignment” of outcome variables to what’s taught in the intervention group. Potential conflicts of interest were also offered as an issue, as evaluators and intervention developers may be the same person or closely connected. As in other areas, this participant also expressed concern about different ways of rigging outcome measurement and analyses to provide the desired findings: “fixing the analysis in advance versus perhaps fixing the outcome variables and fiddling with the analysis until you get the answer that you know is the right one” (P33). Another education researcher echoed the emphasis on attrition, given the nascency of the field in conducting large field experiments, and the importance of attrition to both unbiased estimates of treatment effect, as well as applicability of results to populations other than the study sample (P8). This participant also indicated that evidence for baseline equivalence was rarely reported, and thought it useful information. Another participant who conducts field experiments in educational contexts agreed that power and selection models as they relate to cluster trials in education are a big issue (P28), with standards for reporting intra-class correlation coefficients flagged as an important area to discuss (P7).

The discussant noted the importance of considering the difference between issues for dichotomous data versus continuous data before moving on to views from psychology researchers. Two participants raised the need to focus on non-traditional trial designs (e.g., stepped-wedge designs, N-of-1 studies) for certain social and psychological interventions (P10, P32). The on-going development of a CONSORT Extension for N-of-1 studies<sup>31</sup> was noted in response (P31). Another participant emphasised the fundamental link between encouraging properly reported protocols and then mentioning protocols in subsequent trial publications (P11). One participant then focused on large, complex interventions

embedded within policy systems, discussing the importance of nuanced aspects of trial design, the potential of using administrative data to test theory, and again reporting attrition in cluster trials (P12).

One researcher who specialises in addiction research wondered whether CONSORT-SPI would address interventions where behavioural interventions are ancillary to another intervention, such as nicotine-replacement therapy (P15). This researcher also raised the possibility of presenting the manual for experimental and control conditions as supplements to articles (as is done in an addiction journal), though noting that it may be difficult for journal staff to reliably enforce such a policy. A brief interchange then broke out on motivations of participants to enrol in a trial:

P15: One other thing that I think is very important and cuts across different areas is the study population and the extent to which the population is engaging with an intervention because they want to change their behaviour in some way or they are being approached opportunistically.... perhaps we need an indication of “what is proactive? Was it opportunistic?” Or whatever else you might think.

P6: There’s another aspect to that, and that is whether people are even interested. Because there’s a big difference for example between a depressed individual who is seeking treatment and one whom you screen and say “guess what, I have bad news: you’re depressed, you need treatment.” There are two very different potential outcomes.

This participant then went on to indicate an array of issues to address: diagnosis, selection of primary outcomes, training and qualification of intervention providers, and time and effort spent in data collection as “part of the intervention” (P6).

Following these discussions, a member of the CONSORT Group indicated that most of these comments are common in the medical area and have been discussed at previous consensus meetings, indicating support of this project building on CONSORT rather than being an independent enterprise. This participant also commented that this conversation reveals the need to increase journal word limits, as *Nature* has done in removing word limits for its method section. A criminology researcher asynchronously returned to the issue of enrolment in order to discuss degrees of coercion to recruit participants for trials,

which has myriad implications for interpreting and then generalising trial results for other contexts (P27). Another participant followed-up with various comments, first to the point about journal word limits, drawing attention to increasing critiques of traditional publication practices as well as the move to online publishing (P14). This was followed by recommendations to consider personnel turnover in trials, as well as the unique challenges of measuring systemic outcomes like organisational or community capacity.

Turning to issues in public health, monitoring intervention implementation over time for multi-level interventions was proffered as one area to consider, as well as intervention costs and inequity resultant from the intervention (P25). Akin to other areas, this participant raised concerns about complex outcome measures, governance around potential conflicts of interest, non-traditional trial designs, and using administrative data.

Returning to psychology, another participant stressed the importance of differentiating intervention content from mode of delivery, as well as the application of theory to design and evaluation of the intervention rather than a tokenistic reference in the introduction and discussion sections (P32). Another psychology researcher echoed the importance of detailing the intervention and its theory-base, particularly considering the idea of the provider (i.e., therapist) as part of the intervention (P20). In addition to all of the issues raised above, two researchers in prevention science particularly emphasised the scope of problems related to “cherry-picking” data, particularly given the heterogeneity of outcomes targeted by social and psychological interventions (P12, P24).

To conclude the session, a participant who researches both psychological and pharmacological interventions raised the issue of blinding outcome assessors: e.g., basing practice guidelines on trials of parenting programmes where parents who are extremely invested in their child’s wellbeing are trained to deliver an intervention to their children and then also judge how well the intervention worked (P11). This participant then

developed further issues related to conflicts of interest, indicating that intellectual passion for a particular idea or programme is a much more serious conflict than financial conflicts of interest, yet a difficult one to capture.

### 3.2.6 Session 2.1— Focus of Checklist (Discussant P5)

The discussant began the session with ways of distinguishing social and psychological interventions from medical interventions, such as the mechanisms targeted by the interventions, the techniques used as part of the interventions, the environments in which these trials take place, and latent constructs as outcomes, using the MRC Framework for Complex Interventions as one guide.<sup>32</sup> Two participants wondered why the guideline was called “SPI” instead of “complex interventions” or “complex social interventions” (P25, P5), with another thinking that connotations of “social and psychological” leave out some areas of public health that should be targeted by this guideline (P12). Another thought the proposed distinctions focused more on psychological than social interventions, asking for “physical *and social* environment” to be added to the list of relevant contexts (P14). Another participant suggested that the target or mechanism is the simplest way to distinguish them:

We’re trying to kind of carve out reality in a way that is going to be helpful. But psychological is basically behavioural, emotional, and cognitive. That’s that bit. And then social is at a higher level, so I think having something about the target would define it. So I would have thought that the best bet would be in terms of defining it in terms of the target. (P32)

This then led to brief considerations for changing the title to fit the full scope of targeted areas:

P11: What about public in front of it, or population? Population, social, and psychological interventions.

P30: Population or individual...

P14: Social environment might...

P5: My suggestion is that we go through and think a bit more about these different sorts of interventions and then go back to the labeling it, perhaps?

P29: In labeling, I think its useful to distinguish between the target (what it is to be changed) and the nature of the intervention (is it a chemical? Is it a physical activity? Is it a talk-type of activity?). The setting is another dimension, and I think explicitly separating and thinking through those dimensions may allow for some inclusion and exclusion, but it's at least a three-dimensional table, and it could be more.

The discussant continued with the example of Multi-Systemic Therapy<sup>33</sup> as one intervention targeted by the guideline that fits nicely with the MRC Framework as well as highlights issues to do with implementation, measurement, theory, and context. The discussant then provided an overview of items related to internal validity, external validity, and important information to include in publications (e.g., trial registration number) that would be discussed throughout the meeting. The TIDieR checklist<sup>10</sup> was briefly reviewed once again, and then the floor was opened up for discussion.

Concurrent secular or extraneous events were discussed first, as one participant (P30) sought clarity on the concept. An example was provided:

I was involved in a large study ... where we were trying to get schools to do a curriculum on tobacco control, and during the several years of the study, the state put into place regulations that schools had to do testing for students in order to continue to receive state funding...since our area wasn't covered in the exams, the schools, they were unwilling to do the curriculum because it was a waste of their time. (P14)

Comments about generalisability followed. Harkening back to a previous session, inference and generalisability were proposed to consist of different concepts, with inference as a statistical property and generalisability as a judgment (P14). Another participant emphasised the subjective nature of external validity compared to internal validity, suggesting caution in how it is addressed (P33). Providing details about methods and results rather than requesting judgments on generalisability by the authors was suggested (P30); the value of a flow diagram of participants through the trial to understanding the sample was then underscored (P4). One participant tried to steer the conversation towards neuropsychology and whether interventions in this domain would be

included (P11), but the discussion swiftly returned to external validity, as another participant noted that the concept of external validity is important, but often not its application (P10). To this, a participant who has worked on guideline advisory committees stressed the important of reporting information needed by readers to infer applicability to their populations, as well as explicitly linking to theory (P32). The significance of transparently reporting these features as separate from readers inferring applicability was echoed by two other participants (P14, P18). Given difficulties raised when adaptation is brought into the picture (P22), thinking about how interventions work at population versus individual levels was suggested as one thing to consider for remaining discussions (P25). A richer nomenclature or taxonomy for threats to external validity and scalability was proffered as a helpful avenue of future research, particularly for those working on interventions with greater policy relevance (P12), with another participant noting how difficult it is to objectively describe trial features related to generalisability (P27).

### *3.2.7 Session 2.2— The CONSORT-SPI Delphi Process (Discussant P3)*

This session was relatively brief, as the discussant gave a formal presentation summarising the CONSORT-SPI Delphi process. As this study is discussed elsewhere in the thesis (see Chapter 4), and there was not any participant discussion, it is only worth noting that the discussant walked the participants through the organisation of the draft CONSORT-SPI checklist in order to ensure all participants understood its structure going into voting on checklist items. Namely, items from the Delphi process were mapped onto the CONSORT 2010 checklist, along with their final ranking from the Delphi process, highlighting in green those items that had high support, and those items in red that had indeterminate support. The discussant also reiterated the importance of considering the qualitative feedback during meeting discussions, highlighting key concerns from the

Delphi process about making sure items apply to all social and psychological interventions rather than a subset of them.

### *3.2.8 Session 2.3— Discussion: Results from the Delphi Process (Discussant P13)*

The discussant instructed participants to move from the previous, general discussion to a more focused conversation on the feedback from the Delphi process. One participant familiar with running Delphi processes drew attention to the fact that, though Delphi processes are useful for engaging many people and informing the guideline, there is a common temptation for Delphi participants to “like everything” (P31). This participant pressed that the goal for consensus meeting participants was therefore to use the relative rankings from the Delphi to decide the content for a *minimum* checklist. Another participant experienced in developing guidelines echoed the fantastic engagement in the CONSORT-SPI Delphi process, and that hard, pragmatic choices would need to be made during the consensus meeting, offering the E&E paper as one place to include Delphi items and thus reduce the size of the checklist (P17). If such an approach were to be taken, participants were concerned that the E&E might not be used, and thus, somewhere on the checklist, readers should be instructed not to use the checklist without first looking at the E&E (P11, P20). Another participant reiterated an earlier point that the CONSORT-SPI guideline is really an intervention to improve research reporting, and, considering implementation of an intervention is key during stages of intervention design, the group really needs to sort out how to use the E&E and the checklist in tandem (P32).

A few participants noted that information about trials is often reported across several publications, and the need to clarify whether the scope of this guideline is for the main trial manuscript or all trial manuscripts (P13, P18, P31). One participant suggested developing recommendations for preliminary papers prior to the primary outcome paper, in which

many items related to design could be included (P6). Another returned to the issue of the length of the checklist, stating that a lengthy checklist would be ok if it were put together and presented well, considering the importance of so many items being proposed (P32). Defining social and psychological interventions was again accentuated, as support for items in the Delphi and thus the consensus meeting may depend on one's paradigm and interventions of interest:

There's support ... over some items depending on what paradigm you sit in, whether or not you're more likely to be thinking individual-level interventions vs. population-level interventions. So ... I wonder whether or not we need to just be a bit clearer from the start ... so that you can get a sense of what paradigm that you're in.... Because we could reject some recommendations that are really important for one field in the spirit of efficiency, but without knowing who drove the Delphi responses. (P25)

An experienced guideline developer responded in affirmation, as this illustrates the difficulty of steering a checklist between alternatives from various viewpoints (P1). After noting the large number of items that were recommended by the Delphi process for consideration at this consensus meeting, the participant agreed with the idea of the checklist consisting of minimal, essential standards. Desirable standards that are either not essential or are already subsumed by existing CONSORT 2010 items could consequently be elaborated in the E&E document, as many of the proposed items apply to all trials, not just those on social and psychological interventions. To this, a participant maintained that conceiving of the guideline and E&E as teaching tools should be prominent throughout decision-making at the meeting (P20). Another participant liked the distinction between "desirable" and "essential" items, though worried that a long, narrative E&E may not be effective (P12).

This participant then turned the conversation back to the scope of the guideline. Drawing attention to the large amount of psychology researchers in the Delphi process, as well as the presence of many fields doing complex social interventions, mechanisms were

proposed as the starting point for defining social and psychological interventions:

We're referring [to] interventions where the intervention, the mechanisms of change or targets of outcomes, are social, emotional, cognitive, behavioural in nature. Something like that. That doesn't mean that they might not have physiological outcomes, too. It might not mean other things. But I think some kind of functional analysis of the intervention, the mechanisms of action, and the outcomes, and some description of it that includes ascendant groups. (P12)

The utility of this distinction rather than one by the disciplines targeted was reiterated (P6, P12, P32). With this anchor in mind, a participant with experience as a journal editor underscored a desire for the checklist to be simple and short, and again focus on transparent reporting rather than controlling how researchers conduct trials (P10). Another journal editor related this comment to a desire for the minimal checklist to focus on reporting the intervention in sufficient detail, with the E&E illustrating examples via case studies so that readers can understand what may be useful in their work (P22). In making these tough decisions, another participant made a plea to show deference as appropriate to the collective wisdom of those in the CONSORT Group and others involved in developing guidelines who were in attendance (P33). One such participant suggested that the group could consider: mandatory and optional items, as is done for the PROSPERO<sup>34</sup> registry for systematic reviews; having the E&E and checklist in one paper, as is done for the CONSORT Extension for Non-Pharmacologic Treatments;<sup>17</sup> and making YouTube videos on particularly important items or issues in the CONSORT-SPI checklist (P17). This participant stressed that the ultimate goal is to make sure the guideline is endorsed and used, and for the group to consider implementation throughout the meeting.

### **3.3 Day 2 Meeting Minutes**

#### *3.3.1 Session 3.1— Introduction: Previous Evidence, Analytic Framework, Objectives (Discussant P5)*

This session began with a discussion of conceptual frameworks. One participant

began by arguing that, for social and psychological interventions, reporting mechanisms of action are critical, yet a reference to another document describing this would be adequate (P29). Two participants raised concerns about authors inventing and then reporting analytical frameworks at the end of the study, rather than having the opportunity to report they did not have a framework in the first place (P15, P17), to which the goal of transparently reporting core aspects of these trials was reiterated (P10, P17). Based on the Delphi process results, one participant suggested an item on mechanisms of action in the checklist, with guidance in the E&E that a full conceptual framework or logic model may be helpful for more complex interventions (P3). The term “logic model” then required definition, highlighting the need to document technical terms:

P26: Can I just ask real quick—what’s a logic model?

P11: Yeah can we not use the term “logic model”?

P25: It is a very well-used word, well increasingly well-used, certainly within systematic reviews within the public health group of Cochrane there’s an expectation that a theoretical framework or logic model will be drawn up into the systematic review.

P28: Some people call it “programme theory”.

P32: So I think this is where we need the language all documented, you know, some glossary that these alternatives...

P11: Theoretical rationale.

P29: And “mechanism of action”.

Once defined, logic models were questioned as potentially going beyond an original purpose of CONSORT to promote the reporting of essential elements of trials related to rigour and bias (P6, P11). Another participant argued that a trial in this area *not* explicitly based on some conceptual framework lacks rigour, which should not be narrowly defined in terms of risk of bias (P32). Another offered the rejoinder that the group shouldn’t necessarily push authors to *explain* effects, as the primary purpose of an RCT is *detecting* effects (P26).

Members of the CONSORT Group then asked whether the particular items under discussion required an extension item in the CONSORT-SPI checklist, or whether they were subsumed by existing CONSORT 2010 items and thus could be discussed in the E&E. Strong recommendations in both the E&E as well as editorials accompanying the published guideline documents could be ways to encourage reporting of items, as well as better trial methods (P17, P25). A few participants requested first that the wording of the proposed item on mechanisms of action reflect that trialists may not have a hypothesis for how the intervention works, and the item should involve lay and general language to express this concept (P15, P17, P21, P27, P32). A member of the CONSORT Group reminded participants that the exact wording is not for this forum but for the Project Executive to work on later. With this conceptual change to the item, one participant argued that there isn't a more important issue for this guideline than greater clarity of how the intervention is thought to work and linking this programme theory to other aspects of trial design, such as measurement and choice of outcomes (P12). However, other participants questioned whether these details may be speculative rather than actually tested in the trial (P33), or conversely may have been highly tested in previous literature and thus don't need to take space in the trial manuscript (P24). Another participant provided an example from their own work on how discussing theory in light of a trial's results usefully led to the refinement of that theory (P27). Another followed that such an item would assist with the translation of trial findings to practice and policy (P11).

At this point, the group was instructed how to use the hand-held voting clickers and on the process for voting on the concepts for the checklist. It was clarified that the purpose of tallying votes was to determine whether a relatively high figure for inclusion existed, so as to guide the Project Executive in ruling particular items in or out of the checklist (P18).

The group first voted on "The problem(s) or issue(s) targeted by the intervention(s)",

which received a 60% vote for inclusion. The next item, “Current knowledge about the intervention”, only received a 40% vote for inclusion. The final item for this session, “Whether or how the intervention is hypothesised to work”, received a 73% vote for inclusion. A member of the Project Executive indicated that this item would be perceived to have relatively high support for inclusion in the checklist. One participant suggested that the Project Executive wait until the end of the day to see how oppositional the group was as a whole when judging the relative importance of certain items (P12).

Following these votes, one participant indicated that they would have raised more discussion on the item related to the problem targeted by the intervention had they known the vote would be so low (P24), leading to a discussion on existing CONSORT 2010 items and their explanation in the E&E:

P1: I mean, I accept that, and it is in the text of the E&E paper already. Personally I think the first two [items voted on] are subsumed in what’s already there. That’s my personal view.

P24: They ought to be. People don’t do them.

...

P32: But there [is] some E&E stuff that seems to be words and not in the checklist....

P1: Well, I mean you know, what it currently says is “scientific background and explanation of rationale.” It depends on how you deconstruct that, but we would argue that the reference to prior evidence is part of the scientific background.

P24: But lots of people don’t seem to realise this, especially in a field where people invent their own interventions.

Two participants suggested then that, for future discussions, participants clarify if they are not in favour of an item either because they don’t think it is essential for the checklist or because they think it is already covered in a CONSORT 2010 checklist item (P9, P32).

**Box 1. Summary of Session 3.1 on the Introduction**

“Whether or how the intervention is hypothesised to work” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The CONSORT-SPI E&E should recommend discussion of prior evidence as part of the scientific background and request a full conceptual framework or logic model for particularly complex interventions.

*3.3.2 Session 3.2— Trial Design (Discussant P1)*

To start the session, a member of the Project Executive reminded the group that there would be a few sessions on Day 3 to review decisions made on Day 2. The conversation then first turned towards discussion of potentially including an item referencing the CONSORT Extension for Cluster Trials.<sup>35</sup> A member of the CONSORT Group noted that an online tool was being developed to help authors combine relevant extensions for their particular trial (e.g., a cluster-randomised, social intervention). This participant also mentioned that the Cluster Extension is referenced as a footnote in the CONSORT 2010 Statement, though this group may wish to have an extended item referencing it. The participant also asked the group to consider whether the term “cluster” applies to the place-based intervention mentioned on Day 1. This language issue with the word “cluster” was noted as a common theme in the Delphi process, as was the desire to refrain from mentioning clusters in every item in the CONSORT-SPI checklist, for this would be cumbersome and not always applicable. One participant then suggested that the checklist could indicate which items may need modification for cluster trials (P1). Another participant suggested, as a general principle, that the group should continually consider extended items in the CONSORT for Non-Pharmacologic Treatments extension as well (P17). Another general point related to how to use the Delphi process results in making

decisions, particularly if the group seems to hold a different view than the Delphi consensus:

P18: I mean are we outvoting 300-odd people?

P1: ... I don't think those 380 people were as clued up to exactly what we're trying to do here as we now are in this group. Because with the best [explanation] in the world, it was very hard to explain in the text exactly what was meant by "include". At least, that's my impression.

The group then voted on the item "If the unit of random assignment is not the individual, please refer to CONSORT for Cluster Randomised Trials", which received a 93% vote for inclusion.

The group then discussed an item on stakeholder involvement in the trial. Two participants asked for clarification, as to whether this item involved conflicts of interest (P25) and to which stakeholders this referred (P17). The phrasing of the item suggested that these questions were up for discussion (P3). Two further participants highlighted reporting stakeholder involvement as an important issue to transparently indicate in an article (P12, P29). A few participants then questioned whether "trial design" is the best section to have this item, as opposed to the section on interventions (P7, P22, P24), considering the current section focused on technical aspects of trial design like allocation ratio. Several participants proposed that it have its own category in the checklist if possible, as it did not fit neatly into any other categories (P5, P6, P11, P18). One participant suggested to not frame it as a "conflict of issue" problem, as that might have the iatrogenic effect of preventing transparent reporting, but rather thought it should be framed as a descriptive standard: "It's easier to tell if you just tell them to describe what they did and then you can decide whether it's a conflict or ask them if it's a conflict" (P27). While taken as good suggestions, the discussant recommended that the group first vote on whether it should be included, as a vote of exclusion would make framing irrelevant. The item "Stakeholder involvement in trial design, conduct, and/or analyses (e.g., practitioners,

policy-makers, participant representatives)” received a 73% vote for inclusion.

The session concluded with the discussant allowing participants to raise potential items about sample size if desired, as none were proposed through the Delphi process. One participant sought clarification about the unit of randomisation in the current CONSORT 2010 item (P12). Once explained that the Cluster Extension would cover this (P12), the discussant decided a vote on new items for this section was not necessary.

**Box 2. Summary of Session 3.2 on Trial Design**

“If the unit of random assignment is not the individual, please refer to CONSORT for Cluster Randomised Trials” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

“Stakeholder involvement in trial design, conduct, and/or analyses (e.g., practitioners, policy-makers, participant representatives)” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The CONSORT-SPI checklist should indicate which items may need modification for cluster trials, based on the CONSORT for Cluster Randomised Trials Extension.

The CONSORT-SPI checklist should have a new section for stakeholder involvement.

*3.3.3 Session 3.3— Randomisation Procedures (Discussant P17)*

This session was brief, as no new items were suggested from the Delphi process for this section. The discussant brought up the extended item for sequence generation in the CONSORT Extension for Non-Pharmacologic Treatments: “when applicable, how care providers were added to each trial group”.<sup>17</sup> After clarifying the item referred to how providers were assigned to groups (P12, P24), some participants felt that such an item would need to broaden to other types of interventions than those involving “care providers” (P32), at which point a few participants indicated support of voting on this item.

However, they argued that it may not be appropriate for the sequence generation section given that providers may not be randomly assigned (P17, P20). After the item was rephrased to “When applicable, how intervention providers were assigned to each group”, the item received a 90% vote for inclusion.

**Box 3. Summary of Session 3.3 on Randomisation Procedures**

“When applicable, how intervention providers were assigned to each group” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The CONSORT-SPI E&E should clarify the different types of “intervention providers” across social and psychological interventions.

*3.3.4 Session 3.4—Blinding Procedures (Discussant P3)*

This section began with a brief aside from a member of the Project Executive asking participants to fill out, before the end of the meeting, forms related to dissemination activities down the line. Another member of the Project Executive noted that, while the Executive itself as well as the Delphi process had an equitable gender balance, the Executive was aware of the uneven gender balance at the consensus meeting, and that activities post-meeting will seek to address this imbalance.

To ground the conversation on blinding procedures, the discussant noted that this was a contentious issue in the Delphi process, both in terms of the conceptual applicability of blinding participants and providers of social and psychological interventions, as well as the terminology to be used (e.g., “masking”). One participant found an item on awareness of intervention assignment by providers and participants as speculative (P11). A member of the CONSORT Group then clarified that methods to assess the success of blinding had been removed from the CONSORT 2010 Statement due to methodological concerns about its utility. Another participant suggested that the underlying concern was bias, and that an

item related to blinding in this area might benefit from broadening out (P29). In response, the ability and importance of blinding outcome assessors in social and psychological interventions was affirmed, and suggested such blinding to be considered separately from blinding participants and providers (P11). However, a systematic reviewer indicated the difficulty of maintaining assessor blinding throughout the trial, and thus lamented that trials being appraised in systematic reviews often don't address this issue (P16), which was echoed by a researcher working on parenting programmes (P9).

A criminology researcher expressed concerns about evaluating Stable Unit Treatment Value Assumptions, as those in the criminal justice system may be acutely aware of other intervention groups and respond quite adversely if assigned to a control group (P27). This researcher argued that an item about awareness of intervention assignment might help to appraise whether this issue is relevant in a given trial. To address this issue, a distinction about level of awareness was made:

You have to make clear “[aware] of what”, because I’m not sure that ... people I work with would necessarily think with the way we ... [are] wording it, to say that they did or did not make sure in their consent procedure that the people weren’t aware or were aware that there was another condition and what that condition might be. So for example, when we’re doing an RCT of an intervention, we might say “you’re taking part in a study to evaluate interventions to do X”, and not necessarily say “you’re in the intervention group, and you’re in the control group”. (P15)

It was noted that some disciplines may not have this issue, as ethics committees would require investigators to inform participants of all trial groups (P26). A few participants then requested to have the checklist item be changed from “blinding”, deemed too medical a term, to something like “awareness of assignment” (P7, P11), as well as indicate status of blinding throughout the trial (P16, P26), particularly as it relates to awareness during baseline assessment (P9). A concern was raised about the applicability of these items to community-based interventions, where such awareness would be hard to assess, and consequently mechanisms put in place to reduce biases related to awareness are more

important (P25). This concept was argued to be included in the “and how” clause of the existing CONSORT 2010 item, making deconstruction of this concept in the E&E crucial (P12).

When brought to a vote, the item “awareness of intervention assignment by providers, participants, and assessors throughout the trial” received 57% support for inclusion. This was interpreted to mean that this concept is important but better suited for the E&E rather than the checklist (P12, P16, P18). By comparison, the item “similarities as delivered and received” had only a 23% vote for inclusion, possibly due to overlap with items in the intervention section as well as being addressed by the current CONSORT 2010 item.

#### **Box 4. Summary of Session 3.4 on Blinding Procedures**

Language about “blinding” in the CONSORT-SPI checklist should be replaced with “masking” and “awareness of intervention assignment”.

The CONSORT-SPI E&E should discuss the role of awareness of intervention assignment by providers, participants, and assessors throughout social and psychological intervention trials.

The E&E should affirm the ability to and importance of masking outcome assessors in social and psychological interventions.

#### *3.3.5 Session 3.5— Intervention Design (Discussant P13)*

The discussant drew participants’ attention once again to the new TIDieR checklist, as many items proposed to intervention design related to items in this new checklist.

Another participant involved in the TIDieR project explained the process of developing the guideline, which was very similar to the CONSORT-SPI project plan in following the EQUATOR Network guidance: literature reviews, a Delphi process, and a consensus

meeting. Following this summary, a suggestion for the CONSORT-SPI checklist was to have an item listing the 12 “labels” on the TIDieR checklist that refer to different dimensions of an intervention (including comparator groups), and then referencing TIDieR in the E&E document (P32). Another option proposed was to just reference TIDieR (P13), with an eye towards staying consistent with how the CONSORT Statement itself may incorporate TIDieR (P17).

The conversation then largely revolved around concerns about the applicability of TIDieR to more complex interventions (P12, P25). Firstly, a psychological intervention researcher stressed their desire for the checklist to address provider competence (P20), with another psychology researcher echoing the importance of differential competence amongst providers within the same intervention group as well as between groups (P26). A criminology researcher emphasised the importance of intervention items in the CONSORT-SPI checklist to address the intended duration of an outcome in relation to the intended duration of intervention (P27). The discussion quickly returned to the issue of competence, with debate as to whether it should be considered part of the intervention or a separate issue, which was agreed to be detail that needs to be borne out in the E&E document (P1, P11, P22). Several researchers responded again that this conversation seemed more focused on psychological interventions rather than more complex social interventions (P14, P32), such as those that target infrastructure (P21). To address this, there may need to be a loosening of the items in the TIDieR checklist for CONSORT-SPI and an explanation of their meaning in a box in the E&E (P17):

The social, the complex social interventions: it's not just individuals vs. cluster as the unit of randomisation, it's the unit of intervention, and the unit of intervention is "less tidy" in that, and this [TIDieR] will not capture the key descriptors of the intervention that will allow you to replicate at many of the levels of social intervention that I'm aware of. That doesn't mean this isn't progress, maybe we note this in the box, and we note that future work needs to be done for templates for intervention descriptions for even more complex, kind of systemic level social interventions. (P12)

Several participants indicated support of discussing and possibly even developing a version of TIDieR as it relates to social and environmental types of interventions (P13, P17, P32), for example, how to conceptualise “who provided” the intervention and “how” for an organisational change intervention or policy (P14). One participant responded that the CONSORT 2010 item as it stands is useful in asking authors to report “what happened” to the extent needed for replication, and that perhaps TIDieR should just be referenced and discussed in the E&E (P25). One point in rebuttal was that not including some of the TIDieR labels (e.g, “what” is the intervention) in the checklist will lead authors to continue to describe the intervention more so in terms of its mode of delivery rather than the actual active ingredients (P32).

A vote on including the TIDieR labels in the checklist was proposed, with the understanding that a vote of “Exclude” might mean participants still want to see TIDieR discussed in the E&E. This item received a 60% vote for inclusion. One participant noted that the Delphi items related to the TIDieR checklist received large support, and that TIDieR is likely to be incorporated in CONSORT in some way, so it should be incorporated in CONSORT-SPI for consistency (P13). One participant noted that, while the items ranked in the Delphi process were indeed important, the current CONSORT Item 5 addressed these issues in its general wording, and thus thought description at the level of the Delphi items might be better suited for the E&E (P6). In response, two participants again raised the body of evidence indicating that reporting of interventions is still quite poor, and thus CONSORT Item 5 as it stands may not be having its intended effect (P17, P32). A vote was then proposed to explicitly include reference of TIDieR in the CONSORT-SPI E&E, which received 97% support.

**Box 5. Summary of Session 3.5 on Intervention Design**

The CONSORT-SPI E&E should have a box on the TIDieR Checklist, explain how to use it, and note that future work needs to be done to adapt it to more complex social interventions.

The E&E should address issues of provider competence.

*3.3.6 Session 3.6— Intervention Implementation: Setting, Delivery, and Uptake*

*(Discussant P5)*

After the discussant reviewed the results from the Delphi process and the suggested items, one participant expressed that CONSORT as it stands seems adequate, subsuming much of what was proposed (P17). Another participant responded that demands for contextual information are large in this area, and the group may want to extend CONSORT in particular areas that current CONSORT items do not typically elicit (P18). Particularly important might be the features of context that differentially influence the intended mechanisms of the intervention and comparator, as well as those that impact the ability to implement the intervention (P12). One example provided was the synergistic, interactive effects of intervention components that may be dependent on the delivery system in a trial (P29). A vote was then taken on “important features of intervention setting(s), delivering organisation(s), and wider context”, which received 60% support.

A discussion on actual delivery and uptake of the intervention by providers and participants followed the discussion on context. To start, participants debated whether a distinction between delivery and uptake was necessary:

P11: Delivery and uptake: is that distinction necessary in that item, because that seems a bit...just delivery surely?

P5: No, I don't think so. There's quite a distinction surely between a parenting programme that's got 10 sessions being delivered, and then whether the participants showed up for them.

Once agreed that the distinction was important, a further distinction was clarified between the methods to assess actual delivery and uptake, and then the actual delivery and uptake of interventions themselves (P27). Concern was again raised about the applicability of these concepts of “providers” and “participants” to some complex social interventions, like policy change around smoking in the workplace (P27). As such, these terms were taken out of the conceptualising, leaving the item just to focus on delivery and uptake more generally (P25). A vote was then taken on an item relating to “how and by whom actual delivery and uptake of the intervention was assessed”, which received only a 63% vote for inclusion.

Prior to a vote on the actual delivery and uptake of interventions (rather than methods to assess this), a conversation on tailoring and modification took place. One participant suggested that the item keep to lay language (P27), and then specific terms such as tailoring, modifications, and adaptations as they apply to various disciplines could be discussed in the E&E (P30). One definition of tailoring involved part of the design, delivering slightly different versions of a programme depending on the characteristics of participants (P14). While agreeing the possibility of explicit adaptation built into intervention protocols, another participant thought any item on this topic should be amenable to unplanned or implicit adaptations:

Ideally there would be an explicit adaptation/tailoring protocol built into the intervention. That's rarely done. Oftentimes there is some sort of hand-waving guidance for tailoring/adaptation. That's more often done. And we also see unplanned or perhaps more implicitly developed and guided adaptations. To me, no matter what, we'd like all of that to be documented. I'm not sure that we need to distinguish between the pre-planned vs. the emergent. The point is, let us know what happened and why. (P29)

This point led to a debate about how many papers are needed to report all of the important information for SPI RCTs, as practically all proposed items are important to some degree:

P6: I must say that with respect to every one of these items, I believe that they are all important. What I'm struggling with is that what we're dealing with is the primary outcome paper, and I really think that that level of detail is overwhelming for a primary

outcome paper.... It's almost as if these fields require two papers in order to cover all of the things that are important.

P32: So I really don't agree.... it's not just about saying "here are the data". Its about saying "here are the data, and we'll give you enough information to get to some understanding about what these data mean, and therefore what you can do with the data". And I think that we are just gonna do our collective discipline such a disservice by going down that very narrow—and I think its a biomedical model—of trying to shovel stuff into that, and to then be saying let's chop up these crucial bits of evidence that are required to understand what the data mean into different papers. I just think it's disastrous.

After these comments, the discussion was moved to a vote, at which point the item “extent to which intervention and control conditions were delivered and taken up as planned” received a 97% vote for inclusion.

**Box 6. Summary of Session 3.6 on Intervention Implementation**

“Extent to which intervention and control conditions were delivered and taken up as planned” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The E&E should address issues of tailoring, modifications, and adaptations as they apply to various disciplines.

*3.3.7 Session 4.1— Data Collection and Analytic Plan (Discussant P1)*

The first issue raised concerned psychometric properties of measures, particularly for longitudinal designs that use different measures to assess the same construct (e.g., reading ability in elementary school), and for using a measure across different contexts (e.g., the US and Sub-Saharan Africa) (P12). Another participant then asked to clarify whether “pre-specified” primary outcomes required trial registration (P10), to which the discussant indicated that authors increasingly are publishing trial protocols, which should be considered pre-specification if mentioned there.

Issues related to the length of follow-up of measures were then discussed. One participant suggested that length of follow-up should be addressed earlier when discussing

the trial objectives and the intervention theory of change (P7), while another thought it may not be as pertinent to interventions that aren't expected to work once the intervention is removed (e.g., seatbelts) (P27). From these points, a few participants indicated that perhaps all of the proposed extension items for outcome measures are subsumed by the existing CONSORT 2010 Item 6a:

P21: These all strike me as completely generic issues, and not necessarily ones that are particularly problematic in relation to SPI type interventions, so I wonder whether we should be using our scarce, you know, resource space on the list for this kind of item....

P1: ... I'm much more likely to vote in favour of adding something if I could see that it had a particular relevance for these types of trial, and as you say I would hold my endorsement for those....

P19: I mean I didn't think that they add anything to the existing CONSORT. I would be amazed if anyone read the existing CONSORT statement and think they didn't apply to everything in those ones.

From here, the discussant proposed voting on both amendments to CONSORT 2010 Item 6a: “choice of outcome measures and length of follow-up”, as well as “mode of administration, modifications to pre-existing measures, and differences across groups outcome measurement”. This item received a vote of only 16% include.

Participants then discussed an item on identifying all outcome measures in the trial and if these match the trial protocol. One participant noted that protocols are not required, and the existing CONSORT 2010 Item 6b covers the issue sufficiently, so this extended item would be illogical to include (P12). Another participant suggested the focus then turn to whether trial reports should indicate all outcomes measured in a trial (P13). Someone argued that the classic flaw that the group should target is a trialist changing their primary outcome measure after the trial has begun (P19), to which another participant indicated that issue is addressed by CONSORT 2010 in the checklist and its E&E paper (P17). The discussant then asked the group to consider whether the proposed text facilitates, clarifies, or confuses the existing CONSORT item. A few participants thought it might be useful to determine whether things like process measures are considered outcomes (P12, P26), to

which the discussant proffered that anything after randomisation might be considered an outcome. One participant revisited the utility of having authors report all measures from the trial in the outcome paper (P18), given that a very small proportion of social and psychological interventions RCTs are actually registered (P3, P13). At this point, the discussant suggested a vote, with only 19% in favour of including an item like “identify all outcome measures in the trial, and if these match the trial registration/protocol”.

The session discussion then turned to statistical analyses. One participant strongly argued in favour of an item related to reporting imputation methods, given first-hand experience of an authorial team changing their imputation method across four manuscript submissions for the same trial (P19). Another participant emphasised that methods for handling missing data differ and are more challenging for continuous outcomes, which are prevalent amongst SPI RCTs (P13). The discussant then turned to a proposed item on rationale for statistical methods, indicating a preference to not have items that ask authors to justify their methods, yet a participant responded that they find explanation and providing reasons for analyses beneficial as an editor (P16). Another participant reiterated the importance of transparent reporting of reasons for analyses, though indicated this may be more appropriate for the E&E.

At this point, the discussant suggested the group vote on items in turn. An item for “rationale for analytic methods” received only 29% support for inclusion, whereas an item for “imputation methods for handling missing data and any transformations or changes to data, with reasons” had 68% in favour of inclusion. Lastly, an item related to “whether analyses match the trial registration/protocol” had 61% in favour of inclusion. The discussant took these scores to mean that all of these issues should be discussed in the E&E text, though perhaps the imputation item could be brought to vote again on Day 3.

**Box 7. Summary of Session 4.1 on Data Collection and Analytic Plan**

The E&E should address the psychometric properties of outcome measures, the rationale for analytic methods, and the match between reported outcomes and those listed in the trial protocol.

*3.3.8 Session 4.2— Participant Flow and Recruitment (Discussant P31)*

The first item discussed was eligibility for providers and settings. One participant asked if this was covered by the CONSORT Extension for Cluster Trials, indicating concerns about use of the word “participants” as it connotes individuals as the unit of intervention (P12). Another participant suggested that the group bear in mind the CONSORT Extension for Non-Pharmacologic Treatments, which has an extended item on eligibility for centres and those performing the intervention (P17).

The discussant then turned to the item on rationale for the comparator intervention. While one participant thought it may not fit as an item under participant eligibility criteria (P11), several (P10, P17) indicated that it is an important item in order to understand the information generated by the trial: “by encouraging people to be explicit in the primary report of the study I think can help both reviewers and editors and the readers to focus on what it was that you were, why you think that comparator allows you to answer the question” (P15).

Participants then discussed the item on incentives. This item was seen as important, with incentives taking different forms in social and psychological intervention trials. Examples include payment to attend intervention sessions (P30), coercion to participate in a trial (P11), or rewards that lead to high retention rates (P4, P19). Following these points, debate ensued as to whether this item related to the intervention (P26) or a trial more generally (P13), so as to determine which section of the checklist it might best belong.

The discussant then asked participants to vote on the item “eligibility for providers and settings as applicable”, which received only a 45% vote for inclusion. However, it was then noted that this item was in the CONSORT for Non-Pharmacologic Treatments checklist (P17), and that it received a very high score in the Delphi process (P24), though the wording voted on was different from both of these sources, which may have led to the low scores. One participant particularly noted that, for setting level trials, such an item is crucial, and thus it may need to be reconsidered by the Project Executive. Participants then voted on the “rationale for choice of the control/comparator”, which had a 52% vote for inclusion. The item “incentives offered as part of the trial” was then brought to a vote, and had a 90% vote for inclusion.

The discussion then turned to the CONSORT flow diagram and proposed items related to recruitment and retention. It was noted that reporting the number approached and screened may not always be possible (P16), and even when possible the data may not have been collected (P1). One participant was unclear as to whether information on those approached prior to randomisation was helpful in interpreting the intervention effect estimate (P8), to which a few other participants responded this information could help give a sense of how representative a trial’s sample is to one possible target population (P5, P21, P26). The flow diagram for the Non-Pharmacologic Treatment extension was then discussed, as it has additional items for intervention providers as well (P13). The group then voted on the item “where possible, the number approached, screened, and eligible prior to random assignment, with reasons for dropout (and adjust flow diagram accordingly)”, which received an 81% vote for inclusion.

The group then turned to the proposed item on participants discontinuing the intervention but remaining in the trial. It was clarified this could apply to participants who stopped participating in the intervention but showed up to outcome assessment (P30) or

had available administrative data to use (P12). While echoed as an important concept (P26), a few participants suggested it is not necessary for the checklist but rather is suited for the E&E (P14) and as an part of the example flow diagram (P5). When brought to a vote, only 19% favoured inclusion of “number of participants who discontinued the intervention but remained in the trial” in the CONSORT-SPI checklist, though the discussant clarified the sense of the room was to amend it into the flow diagram.

The next item related to dates and timings of study procedures by trial arm. One participant noted that such information is helpful to articulate for interventions that involve many components and strategies, as it may be difficult otherwise to tell when an intervention has actually started (P25). Another participant noted that this item could help clarify previous concerns about whether baseline assessment occurred pre- or post-randomisation, as well as problems determining whether follow-up is defined relative to the beginning or end of intervention delivery (P3). Many participants then indicated support for including time-points in the flow diagram (P11, P12, P25, P26). For the checklist, though, an item on “dates and timings of study procedures by trial arm” received only a 65% vote for inclusion.

The final proposed item for this session involved why the trial ended or was stopped, namely as to whether a trial may still be on going. One participant noted the importance of this depends on the definition of “on going”, such as whether extended follow-up after a primary endpoint mean a trial is still on going (P1). Another participant thought the item linked to sustainability, which is an increasingly important area for SPIs to address (P25). A participant who does field experiments in unstable contexts thought it might also relate to trials that had not ended but perhaps were interrupted for a period of time, for example by conflict in a region (P12). A potential concern though was that an item asking participants to indicate if a trial is on going might encourage trialists to report trials at an

interim stage “under the influence” of the preliminary data (P16). When brought to a vote, the checklist item “indicate other results from trial, with references if available” only received a 55% vote for inclusion.

**Box 8. Summary of Session 4.2 on Participant Flow and Recruitment**

“Incentives offered as part of the trial” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

“Where possible, the number approached, screened, and eligible prior to random assignment, with reasons for dropout (and adjust flow diagram accordingly)” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The CONSORT-SPI flow diagram should ask for the number of participants per group who discontinued the intervention but remained in the trial.

The E&E should discuss the reporting of dates and timings of study procedures by trial arm.

*3.3.9 Session 4.3— Baseline Data (Discussant P3)*

The session began with a discussion about the proposed item relating to reporting socioeconomic data. One participant thought that any such item should have a “where applicable” clause, as not all social and psychological intervention RCTs may collect this data (P6). As baseline variables of interest change over time and vary across disciplines, another participant suggested a general recommendation for characteristics important to understanding the intervention effect (P29). Another participant asked the group to focus on the aspect of the proposed item recommending socioeconomic data, as the impact of interventions across different socioeconomic groups is an increasingly established objective of this area of research and also has a PRISMA extension (P25). Some

participants asked for clarification on what is meant by “socioeconomic”:

P12: And socioeconomic includes I guess, demographic includes cultural?

P25: Yeah, that's why PROGRESS is there, because its place, race, occupation, gender, religion, education, social support, and then a few others.

P3: And then "plus".

P25: You choose them in turn. They have to be relevant to the intervention.

One participant working in systematic reviews argued that socioeconomic data at the aggregate level of a trial does not provide the individual variance that is really needed to assess equity issues (P13), though some participants argued that the trial level data, though imperfect, is still useful (P7, P15, P25, P33). Another participant emphasised that, at this point, the discussions continually seem to assume individuals as participants, and thus something fundamental to the checklist wording and structure needs to be addressed so that the resultant guidance could apply to all social and psychological interventions (P12). With this issue noted, the group then discussed the need for the E&E to address what “socioeconomic” might mean if included as a checklist item so as not to define it restrictively (P22), and to perhaps look to remove “clinical” from the CONSORT-SPI version of CONSORT 2010 Item 15 (P3). When brought to a vote, the item “include socioeconomic variables where applicable” received a 74% vote for inclusion.

The session concluded with a discussion on comparing study completers to the original sample at baseline. A few participants first sought to clarify that “completers” meant those providing outcome data rather than those completing treatment (P6, P19, P24). One participant raised concern that studies are not powered to detect significant differences between those who provide baseline data and those providing follow-up data, misleading many trialists to claim that data completers are not different from drop-outs (P13). While another participant agreed with this concern (P8), a second noted the prevalence of this practice amongst SPI trialists (P24). An item on “differences between those who provide

outcome data at follow-up and original sample per trial arm” only received a 32% vote for inclusion.

**Box 9. Summary of Session 4.3 on Baseline Data**

“Include socioeconomic variables where applicable” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The E&E should discuss the PROGRESS-Plus framework as it relates to equity data.

*3.3.10 Session 4.4— Results: Outcomes and Estimation (Discussant P17)*

At the start of the session, the discussant proposed a vote on whether any changes were needed, given the symmetry between items in this section and those in a previous session on data collection and analysis. After confusion in trying to remember what was discussed in that earlier session, one participant noted that unique concepts were being proposed here (P11), so the discussant opened the floor for comments. To distinguish it from the existing CONSORT item on sample size calculations, the item on power calculations required clarification to mean determining whether each analysis, given losses at follow-up, has the power to detect the originally postulated effect size (P3, P6, P11). At this point, a few participants recommended focusing on power calculations only for primary outcomes (P15, P17). The discussant suggested a vote; an item on “power to detect postulated effect for each analysis of primary outcomes” only received a 32% vote for inclusion.

Turning to the other proposed items on outcome data, one participant argued for explicitly recommending that authors provide summary data by trial group for every outcome and time point (i.e., from baseline to last follow-up), due to first hand experience of trying to conduct meta-analyses but frequently encountering obscure or incomplete reporting of outcome data (P24). A few participants indicated that this is a general problem

for all trials and is already addressed in the CONSORT 2010 E&E document (P1, P16), though it could possibly be proposed for the next version of the CONSORT checklist (P18), as well as explicitly targeted by a box in the CONSORT-SPI E&E (P17).

The group then discussed an item on indicating any other results from the trial. One participant noted that this issue might particularly apply to process evaluations, which are often published as a separate manuscript and need to be linked to any outcomes papers on a trial (P3). Participants responded that the other item on the availability of trial data could better address issues of selective reporting, linking data from multiple publications, and replication (P9, P17). For example, some trials may have dozens of publications (P1), and some journals resist self-citations by authors (P18). One participant raised ethical and resource concerns of requesting trialists to make their data available (P12), to which another suggested that the recommendation be for a data sharing agreement, or indicating the availability of trial data (P10). Several participants noted the movement of intervention research in general down this route (P7, P19, P26). Both items were brought to a vote: the item to “indicate other results from trial, with references if available” received a 55% vote for inclusion, whereas the item to “indicate availability of trial data” received a 77% vote for inclusion in the checklist.

**Box 10. Summary of Session 4.4 on Results: Outcomes and Estimation**

“Indicate availability of trial data” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The E&E should discuss the utility of providing summary data by trial group for every outcome and time point (i.e., from baseline to last follow-up) as well as linking all documents reporting data from the same trial.

### 3.3.11 Session 4.5— Adverse Events (Discussant P13)

This session briefly reviewed a proposed item on adverse psychological events or unanticipated social disadvantage as an extension to the existing CONSORT item on harms. The discussant noted qualitative feedback from the Delphi indicating that an explanation in the E&E detailing how adverse and unintended events should be conceptualised for social and psychological intervention RCTs may be a more fruitful avenue than adding an item to the checklist. One participant noted concerns that harms are over-weighted compared to other outcomes, which typically require pre-specification and rigorous measurement procedures (P10). This problem could be further exacerbated by trials that have numerous variables amendable to fishing (P15). The discussant asked for anyone in favour of the extended item to make a case for it; with no response, the item was brought to a vote, with only 6.5% in favour of including it in the checklist.

#### **Box 11. Summary of Session 4.5 on Adverse Events**

The E&E should detail how adverse and unintended events might be conceptualised for social and psychological intervention RCTs.

### 3.3.12 Session 4.6— Discussion Section: Interpretation, Generalisability, Limitations (Discussant P5)

This session began with consideration of an item about limitations related to heterogeneity and intervention implementation. While one participant thought this issue was already covered in the CONSORT checklist (P20), another thought the item lumped two concepts together, and limitations related to intervention implementation may be worth flagging on its own (P25). Another participant agreed implementation needs to be highlighted, but in the E&E rather than the checklist (P7). As no further support for the item was provided, the discussant called for a vote, and only 19% of participants voted to

include an additional item on limitations in the CONSORT-SPI checklist.

The next item under consideration related to generalisability to particular populations and settings. One participant noted recent work to improve formal theories about generalisability, arguing that the concept of generalisability only makes sense when a particular inference population serves as a reference when asking whether the same intervention effect would occur there, and even then generalisations do not often turn out to be accurate (P33). Another participant suggested not to extend the CONSORT item on generalisability given its utility as currently worded, and because the reader is in a better position than the author to make such judgments anyway (P32). Another participant voiced further concerns about the subjective, speculative nature of most generalisability assessments, though guidance on this could help the field to use more objective, established ways of making these assessments (P10). However, asking trialists to learn a whole new skill set was argued by another participant as a barrier to using CONSORT-SPI (P16). A vote was proposed for the item “population(s) and setting(s) which the results may apply and those which they may not”, which had only a 7% vote for inclusion.

The final part of this session involved an item on implications for future research. One participant recommended that the item also involve policy implications, particularly encouraging authors to discuss the totality of evidence for the intervention being studied when they do so (P24). Participants were asked to consider this item on implications to also involve policy and practice. An item on trial implications received only a 23% vote for inclusion.

**Box 12. Summary of Session 4.6 on the Discussion Section**

The E&E should discuss limitations related to intervention implementation, the inference population for generalisations, and potential implications of social and psychological intervention trials.

3.3.13 Session 4.7— “Other” Information: CoI’s, Ethics, Trial Registration (Discussant P3)

This session began with a discussion of whether participants should indicate in the paper whether a trial was prospectively or retrospectively registered, if registered at all. One participant noted that registries already indicate this information, which therefore shouldn’t be a requirement in CONSORT-SPI (P13). In contrast, a few other participants noted it may be helpful to have this information explicitly noted in the paper rather than trust the reader to go find it elsewhere, as well as advancing trialists in this area to register their trials no matter the status of the study (P3, P12). In an effort to minimise the checklist, another participant proposed that the CONSORT 2010 E&E text already covers this issue, so the CONSORT-SPI E&E could do the same (P16). When brought to a vote, an item like “if not registered, reasons why” or “indicate if registration is a priori” had only a 19% vote for inclusion in the checklist.

The next item related to reporting where any other information about the trial’s methods, outcomes, and interventions could be accessed. One participant noted that materials about the intervention could include video instructions in addition to manuals (P17), to which a participant suggested separating the item into two: one on information about the intervention, and one about trial methods and outcomes (P6). An item on manuals and other related materials was suggested to fit well with CONSORT Item 5 on interventions, if voted for inclusion in the CONSORT-SPI checklist (P1). When brought to a vote, a separate item on “where the intervention manual and other informational materials can be accessed” had a 90% vote for inclusion. Regarding the second item on other information about the trial, one participant noted this now overlapped significantly with an item previously voted on in the section on results (P18). The item on “where other information about the trial’s methods and outcomes can be accessed” only yielded a 45%

vote for inclusion.

The last two items involved potential conflicts of interest. The first related to involvement of the intervention developer in the trial itself, which several participants immediately indicated as very important (P11, P30). Participants suggested it could also include sponsors of the intervention in the setting in which it is being evaluated (P21), as well as the relationship between the developer and the researchers more generally—for example, if the researchers are former students of the intervention developer (P30). One participant raised the concern that, while this information should be in the report, this issue seemed more a policy for journals than for CONSORT to confront (P10). A member of the CONSORT Group affirmed that is why ethical considerations are not considered in CONSORT guidelines (P17). However, a few participants argued that an item on potential conflicts of interest was actually an issue related to bias and interpreting the effect estimate:

P15: I think it isn't just an ethical issue. I think it is, the conflict of interest is as much a potential source of bias as anything else we look at, and so it is critical for interpretation of a study. So I think it does belong here, and I also think that you're right, the journal, good journals would require it. But they don't at the moment, and I think if it goes into CONSORT, into this extension, it will drive practice.

P9: I've looked at it. I think there is a bad, roughly, currently in social science interventions, about 5% of publications have a conflict of interest statement. So it's very, very rare. It's very important to have because, when you look at it, in about 60-70% of RCTs, there is a conflict of interest. And that's not an indirect conflict of interest, but some material conflict of interest. Now the material conflict of interest is not usually the funding of the study, but it's usually either that somebody associated with the study is involved in an organisation that benefits from the dissemination of the intervention, and that's something that's really important to be transparent about.... It doesn't have to be the programme developer.... There is other variance, for example, being on the consulting board of some prevention programmes.

While agreeing that such involvement could be linked to effect estimates, two participants asked the group to remain open to the possibility that it is due to better implementation of the intervention, which is still important to know for scalability reasons (P4, P20). One participant went back to the suggestion that this issue is more for editors to solve, like the

International Committee of Medical Journal Editors does already (P1). In reply, another participant indicated that editors of journals that publish social and psychological intervention RCTs are not as cohesive a group as medical journal editors, and CONSORT is the best way to get these journals to require declarations of interest (P30). When brought to a vote, an item on “any involvement of the intervention developer in the design, conduct, analysis, and reporting of the trial” received an 81% vote for inclusion.

Coming out of this discussion was the proposal for an additional item on allegiance by providers to the intervention being delivered. One participant asked what evidence exists that this item would be important to report because of risk of bias concerns or current poor reporting (P17). Another participant responded that data indicating intervention allegiance is not a causal variable in improving intervention outcomes (P19). Another retorted that, in their field, such data doesn't exist because it is difficult to define, though perhaps putting it in the checklist could bring attention to the issue (P15). It was then suggested that the minimal checklist space might be better used to require transparent reporting of all the data needed to appraise a study instead (P16). An item on “allegiance to the intervention” received only a 23% vote for inclusion.

The last item for the day involved conflicts of interest more generally. A member of the CONSORT Group explained that conflicts of interest are not in the CONSORT checklist because such conflicts are not specific to trials, and they furthermore are an issue for journals to solve with their own policies. This comment led to a quick discussion about the applicability of a journal policy solution for social and psychological intervention RCTs before the vote:

P19: Many journals have their own conflict of interest form you have to fill in to publish whatever.

P24: Many don't.

P19: Don't they?

P24: No, many don't.

P9: It's a different situation in the social sciences.

P19: We publish in different journals.

P12: It's because the social sciences don't have any money involved.

P9: No, that's not true.

P14: That's not true anymore.

P12: I was imagining...

P19: But I think it goes way beyond trials, and it is a policy of the journals to enforce.

When brought to a vote, an item on “any potential conflicts of interest and how they were handled” received only a 19% vote for inclusion.

#### **Box 13. Summary of Session 4.7 on Other Information**

“Where the intervention manual and other informational materials can be accessed” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

“Any involvement of the intervention developer in the design, conduct, analysis, and reporting of the trial” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

The E&E should discuss allegiance by providers to the intervention being delivered.

### ***3.4 Day 3 Meeting Minutes***

#### ***3.4.1 Session 5.1— Review of Day 2 and Aims of Day 3 (Discussant P13)***

This brief session introduced participants to the purpose of Day 3 conversations. Namely, the Project Executive hoped that the consensus meeting group could go over the draft checklist based on votes from the previous day to make sure that “we’ve got the right items in it” (P13). Once these were decided, the group would then discuss whether anything needed to be amended for standards on the title and abstract. The day would then

end with a few sessions related to writing guideline documents, disseminating them to various stakeholders, and seeking proper implementation of the CONSORT-SPI reporting items.

#### *3.4.2 Session 5.2— Tabled Issues from Day 2 (Discussant P3)*

The discussant informed the group that the purpose of this brief session was to ensure that all participants had the chance to raise any issues that hadn't been discussed or didn't seem to be on the agenda for the final day. One participant requested a discussion on the conflict of interest item again, as conversation on this item had been cut short the day before, potentially explaining why this item received such a low vote for inclusion in the checklist (P24). The discussant informed the group this would be considered in part of the next session. Another participant then asked if the group should still refer to the Delphi results when finalising the guideline, in light of decisions the day before that sometimes were at odds with results from the Delphi process (P32). Another participant replied that nothing had really been excluded from the CONSORT-SPI guidance, but rather many items had been more appropriately placed in the E&E to keep the checklist at a minimum. The discussant agreed to have the Delphi results as well as votes from the day prior on hand in case any participants wanted to reference them. The last tabled issue raised was a revisiting of the item for harms (P27), which was also slated for the next session.

#### *3.4.3 Session 5.3—Review and Discussion of Recommended Items (Discussant P5)*

This session involved a review of the draft checklist prepared by the Project Executive after Day 2 sessions had concluded (see Appendix Q). The discussant led the group chronologically through the checklist, flagging issues that the Project Executive noted or asking participants to raise their own concerns or comments. First, a few

participants noted that the Project Executive needed to come up with a better wording for the extended item 2a, “whether or how the intervention is hypothesised to work”, as the wording could be confused to mean whether the investigators think the intervention will work. Another participant asked the group to be mindful through this discussion of the rationale behind including items, particularly if the Non-Pharmacologic Treatment extension (given its similarity to this guideline) did not include the item (P17). One participant argued that this particular item on the intervention’s mechanism had theoretical and empirical importance:

The theoretical reason is that, in this domain, there is a plethora of interventions which have been based on implicit, informal, common sense if you like, understanding or previous practice, and when you’re coming to review those and try to make sense of those, you find it becomes very difficult. The empirical one is precisely that, is that in systematic reviews of evaluations of interventions in this area, there are many cases where you think they probably did use a formal theory, and you’re not quite sure how they used it, and there simply isn’t enough information to say. So I think whatever the situation is with regard to other non-pharmacological interventions, surgical presumably, or other interventions, maybe that issue doesn’t apply, but it certainly does in this area. (P15)

This explanation was judged by the previous participant (P17) as the desired level of rationale for an item that should also be captured in the E&E.

The next checklist item addressed was the desire to reference the Cluster Extension as part of Item 3 in the CONSORT-SPI checklist. Several participants indicated that it should remain in the checklist, rather than as a footnote, given common use of cluster RCTs in this area (P5, P17, P33). This discussion on the Cluster Extension item was followed by a debate as to where the new item for stakeholder involvement in the trial should be located in the checklist. Suggestions included the “Other Information” section, because the stakeholder item crossed several sections of a manuscript (P11); trial design, due to the prominence it would have in this section compared to “other” information (P21); and the introduction section, as this concept is less about technical details and more about helping the reader understand how the intervention was put together (P7). Another

suggestion with some traction was to create a new section called “Stakeholder Involvement”, as it doesn’t fit into any of the other sections without its scope or importance being restricted (P18).

The discussion then moved to the item on participant eligibility, where the Non-Pharmacologic Treatment Extension item on eligibility of centres and care providers was re-raised, given concerns the day before that wording rather than the concept was the reason votes did not reach a critical threshold (P12). It was first noted that the E&E should have a brief section defining “participants” broadly as the “unit of intervention”, so that clusters and places could be included in this concept (P33). Then, when the item was rephrased to “when applicable, eligibility criteria for settings and those delivering the interventions”, the discussant confirmed with the group that they would now be happy with including this item.

As the discussion turned to the section on interventions, there was another plea for the checklist to clearly recommend that authors describe the active ingredients of the intervention (P25, P32). Moreover, those working on “higher-level” interventions reiterated the need for any discussion of TIDieR in the E&E to encourage those working on the “more social, population, environmental based interventions to participate in actively adapting it to their point of view” (P12). After these were agreed, the discussant noted that no changes were made to outcomes, sample size, randomisation, allocation concealment, sequence generation, implementation, and blinding.

When the discussion reached the item on statistics, the issue of handling missing data was again raised due to its vote marginally missing a critical threshold for inclusion. One participant suggested wording for an item from another guideline: “describe how missing data were handled (e.g., complete case analysis, simple imputation, multiple imputation), with details of any imputation method” (P1). Several participants indicated they would be

interested in re-voting on such an item (P12, P17, P19). This time, an item on the concept “statistical approaches for missing data are explicitly stated” received an 84% vote for inclusion in the checklist. A few participants suggested that an item on data transformation be brought to a vote as well (P15, P30), though it ultimately was not, based on the understanding that CONSORT Item 6 as it stands addresses this issue by asking participants to fully define the outcomes in the dataset, and that this point would instead be addressed in the CONSORT-SPI E&E (P13).

The group then discussed the flow diagram. However, contrasted with the Day 2 session on this topic, for this session each participant had a copy of the flow diagram from the Non-Pharmacologic Treatments Extension. The discussant went over this flow diagram on the projector, showing how it includes information about those discontinuing the intervention but remaining in the trial. Several participants agreed it would be useful to have this information in the flow diagram, as well as providing potential examples for cluster or place-based trials (P12, P14, P15, P16).

The group then discussed the new item on socioeconomic data in the baseline data section. Keeping the word “clinical” in the standard CONSORT checklist item here was not seen to be problematic by some participants (P19, P21, P24), though it was suggested—even by those same participants—that more general wording would not lose the concept while also potentially avoiding negative reactions from trialists targeted by the guideline who see “clinical” as medical terminology (P6, P19). As this item again raised the issue of applicability to cluster trials, it was suggested here that an appendix of the CONSORT-SPI extension may include a modified version that incorporates elements from the CONSORT extension for cluster trials (P12, P13).

The new item on the availability of trial data was then discussed. Participants were happy to add this item to the checklist but wanted to match it with a different CONSORT

2010 checklist item than Item 18. One participant then made another case for possibly extending the harms item given recent experience with a journal not accepting data on harms that were not specified in the protocol; in the end, this participant agreed to defer to the group vote and expand on this issue in the E&E. The group quickly moved to the item on the intervention manual and other materials on intervention delivery, which was recommended to be moved to the section on interventions. However, no one proposed a re-vote on the item for referencing “sibling documents” that provide information on the trial’s methods and outcomes. It was noted that an item on referencing the protocol was already in the CONSORT checklist.

This discussion finished with a re-visit of the conflicts of interest item, as many thought this important for the group to fully discuss given that conversation was rushed at the end of Day 2. It was first suggested that the item on stakeholder involvement be distinguished from the item on potential conflicts of interest, as the former is not a conflict of interest but rather about engagement and improving the utility of trials (P7). Furthermore, it was recommended to reconceptualise “conflicts” of interest to “declarations” of interest so as to better encourage people to report relevant information (P25, P30). Some participants felt that such an item would be straying from the original intention of CONSORT (P10). However, it was pointed out that CONSORT Item 25 is called “funding” though really is about conflicts of interest, and that the more important conflicts of interest for SPI RCTs may not be the same as financial conflicts of interest common in pharmaceutical interventions (P15). When brought to a vote, an item on “declaration of any other potential interests” received a vote 81% for inclusion in the CONSORT-SPI checklist.

**Box 14. Summary of Session 5.3: Review and Discussion of Recommended Items**

“Describe how missing data were handled (e.g., complete case analysis, simple imputation, multiple imputation), with details of any imputation method” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

“Declaration of any other potential interests” was voted for inclusion as an additional item in the CONSORT-SPI checklist.

“When applicable, eligibility criteria for settings and those delivering the interventions” was agreed to be included as an additional item in the CONSORT-SPI checklist.

Both checklist items and E&E descriptions for intervention items should clearly recommend that authors describe the active ingredients of the intervention.

Checklist Item 15 on baseline data should use general wording so as to avoid reaction from trialists targeted by the guideline who see “clinical” as medical terminology.

The E&E should discuss the theoretical and empirical importance of specifying how the intervention is hypothesised to work.

The E&E should have a brief section defining “participants” broadly as the “unit of intervention”, so that clusters and places could be included in this concept.

The E&E should include information about data transformation for Item 6 on fully defining outcomes in the dataset.

An appendix of the CONSORT-SPI extension may include a modified checklist that incorporates elements from the CONSORT for Cluster Randomised Trials extension.

#### *3.4.4 Session 5.4— Finalise CONSORT-SPI Checklist (Discussant P13)*

As the previous session also involved finalising the CONSORT-SPI checklist (see Appendix R), this session was used to focus on the appropriateness of the CONSORT Extension title “Social and Psychological Interventions”, due to participant requests that

the title might need to change in order to better engage those working on interventions at a structural or environmental level. One participant suggested adding “systems” to the title, as it captures this theme (P14). Some participants asked why the descriptors “complex” or “complex social” weren’t used (P16), to which a member of the Project Executive explained that the group thought many medical interventions are complex as well, and that many researchers targeted by the guideline may not define their interventions as complex or “see themselves” in the title. A disciplinary approach was offered, with the guideline defined by the disciplines targeted (P12), though when this was tried at early stages of this project, researchers from certain disciplines felt “left out”, and adding disciplinary names to a title became unmanageable (P13). “Environmental” was offered as an additional element to add to the title (P25), though this term had originally been in the title of the project, and peer-reviewers thought that it could lead to confusions with interventions aimed at issues like global warming and forestry (P13).

One participant harkened back to discussions early at the meeting about how these interventions are being defined: if by the mechanism, then something like “environmental” may be applicable, as these interventions work through altering environments (P32). To address this issue of scope, one participant suggested a box in the E&E that explicitly delineates what the guideline developers expect to be targeted by this guideline, so that readers can assess whether to use it (P31). To maintain the momentum behind the project, another participant agreed with this approach and suggested keeping the “CONSORT-SPI” branding, while discussing in the guideline the intended scope. Still, many participants suggested adding at least one modifier to the title before publication (P12, P16).

It was suggested that the Project Executive take on this issue and consult with researchers who work on these “higher-level” interventions, before coming back to the consensus meeting group with recommendations (P5). One participant indicated interest in

voting to see how many participants thought a name change was necessary (P13). While some thought this suggestion to be helpful in advising the group, others thought it missed the point of engaging a small yet significant community in the project. When brought to a vote, only 52% of participants thought the title should remain the same, suggesting that the Project Executive should consider amending the title of the guideline.

**Box 15. Summary of Session 5.4: Finalise CONSORT-SPI Checklist**

The E&E should include a box or table that explicitly delineates the remit of the CONSORT-SPI guidelines.

*3.4.5 Session 5.5— Items for Title and Abstract (Discussant P31)*

The final session on checklist content involved voting on whether to extend items related to the title and abstract. The discussion started off with an argument for extending CONSORT Item 1a to also request that the title include the population and intervention under study in the trial (P24). Another participant responded that, for certain complex interventions, fitting these details in the title may expand beyond what journals allow for character counts (P6). Another echoed that journals like shorter titles, and that providing these details in the abstract would ensure they come up in a literature search (P17). To support the idea for including population and intervention, one researcher thought it odd to report the method in the title but not the subject of the study (P32). Another systematic reviewer supported the idea, as seeing these details in the title could expedite screening (P7). The counter argument revolved around prescriptiveness:

I tend to be anti-prescriptive for a lot of different reasons, one of which is, it frequently doesn't work very well when you don't have control over it, number one. And number two, if you're going to be prescriptive, I think you have to have a very strong rationale for why you're doing it, and I'm not sure that the rationale is there yet for why you want to be so prescriptive. (P14)

When put to a vote, an item in the checklist on reporting the “intervention and

population/problem in title” received only a 52% vote for inclusion. As a result, it was suggested and seconded that this should still be recommended in the E&E, using an example title where these details are reported in addition to randomisation (P1, P32).

The session concluded with a discussion on the abstract. The discussant overviewed the CONSORT for Abstracts Extension<sup>11</sup> and then asked participants if they thought any information was missing from this, or needed explicit highlighting in the CONSORT-SPI checklist. Of the concepts being proposed from the Delphi, time of follow-up was noted as the only one that was not in CONSORT for Abstracts (P18). As no other new concepts were raised, the discussant suggested that the E&E may provide a box with the CONSORT for Abstracts checklist, as well as an explanation of the items within it that is geared towards social and psychological intervention RCTs.

**Box 16. Summary of Session 5.5 on Items for Title and Abstract**

The E&E should recommend that the title include the population and intervention under study in the trial when character limits allow.

The E&E should provide a box with the CONSORT for Abstracts checklist, as well as an explanation of the items within it that is geared towards social and psychological intervention trials.

*3.4.6 Session 5.6— Discussion on Checklist Wording (Discussant P5)*

Originally, the Project Executive planned for a session on checklist wording after the draft checklist had been decided. Because taxonomy and terminology had been discussed so much across all previous sessions, the discussant suggested the group forgo this session to ensure sufficient time to discuss dissemination and implementation of the guideline, and the group agreed.

## **4. Discussion**

### ***4.1 Overall Findings***

The CONSORT-SPI consensus meeting led to the inclusion of 14 additional items for the guideline checklist. The number of additional items in the proposed CONSORT-SPI checklist is similar to other CONSORT Extensions, as many widely used guideline extensions involve anywhere from 10-20 extended items. Compared to other guidelines, however, a thorough, narrative analysis of discussions at the consensus meeting have been made more explicit and transparent for readers to understand and appraise decisions made about guideline content and structure. During the consensus meeting, participants provided rich views on the importance of items to include both in the checklist as well as further discuss in the E&E document. Furthermore, as a member of the CONSORT “family” of guidelines, the CONSORT-SPI Extension will make explicit references to many extensions, including the extensions for abstracts, cluster trials, non-pharmacologic treatments, and harms.

Several general topics were consistently broached during meeting discussions. One important topic was the need to use non-technical terms for the concepts that are targeted by checklist items. This practise would help further ensure the use and endorsement of the checklist across various disciplines. On a related note, many participants noted issues associated with language and taxonomy across targeted disciplines. Researchers from different intellectual backgrounds may use different terms for the same concept, or the same term for slightly different concepts. As such, a glossary defining terms as they are used in CONSORT-SPI guidance could help to provide a shared language for this interdisciplinary area. Another general topic of the consensus meeting was the need for a clearer definition of what social and psychological interventions should entail for this guideline. Participants generally seemed to settle on defining social and psychological

interventions as those interventions that target mechanisms involving psychological phenomena and social processes. As such, both the title and the content of the guideline would likely be relevant to all interventions in this area. Lastly, consensus meeting participants decided to relocate many proposed Delphi items on important concepts from the checklist to detailed discussion in the E&E, in order to keep the checklist minimal. For example, the new TIDieR checklist<sup>10</sup> will be incorporated into the E&E document for CONSORT-SPI. As such, a clear and comprehensive E&E document is likely to be paramount to the impact of the CONSORT-SPI checklist.

#### ***4.2 Strengths and Limitations of the Current Study***

There are several strengths of the current study. Particular strengths of the meeting as noted by participants and the Project Executive include efficient facilitation of discussion, rich qualitative feedback, and anonymously recorded votes for each item. In addition, while facing a brief difficulty at the start of the meeting, clickers proved useful. It is worth considering, though, that rankings may be reflective of the oppositional (or not) nature of participants rather than the importance of items. As such, solely using such quantitative data does not fully capture reasons behind decisions, supporting the utility of the narrative analysis of participant discussion. Furthermore, there is the possibility that participant fatigue could have set in towards the end of a full day or of the meeting itself, as indicated by the quick vote-down of an item on conflict of interest at the end of Day 2 that was then overturned on Day 3. As such, when drafting and editing the final guideline, the Project Executive should continue to confer with the wider group of consensus meeting participants to see if their views have changed at all post-meeting about checklist items, checklist item wording, guidance in the E&E, the title of the guideline, and how best to disseminate and implement the guideline.

Akin to the Delphi process, one potential drawback of this study is the high proportion of participants with a background in psychology and from those who reside in the US or the UK. Discussion and votes by these participants may be motivated by concerns for these areas of research, which would cause problems with the use of majority-votes considering that the guideline is looking to equally address the concerns of all stakeholder groups represented. Nonetheless, both psychology as well as the US and UK research networks do represent a significantly large amount of RCTs in this area, serving as part of the rationale for sampling more participants with these demographics. There was also a greater acceptance rate to attend the meeting from participants in these disciplines. However, differing from the Delphi process, consensus meeting participants were also disproportionately male, with only 20% in attendance being female. The Project Executive is seeking to address this imbalance in gender for future stages of the project.

It is also worth noting that this chapter represents a first draft based on the DPhil candidate's interpretation of meeting discussions. As part of the on-going project, this draft will be sent to consensus meeting participants for validation. However, the current analysis hasn't been double-checked by another coder, it hasn't yet been sent back to participants for their endorsement of interpretations, and there has not been any formal feedback mechanisms after the consensus meeting. This analysis also did not focus on the context of discussion and particular types of informal discourse. While such analyses may have yielded interesting insights about the content discussed, they were seen as beyond the purview of the purpose of this manuscript. Subjective decisions during coding, interpretation, and reporting consensus meeting conversations could not be avoided given the interpretative methods employed in the above representational, narrative analysis. There is an inevitable gap between the selective representation of the meeting above and the full experience of attending the consensus meeting, although the DPhil candidate did

have a lead role in organising and running the meeting.<sup>19</sup> Moreover, this analysis is based off audio-recordings, and thus the content of the conversation has been captured in a way that is less susceptible to bias than if based only on notes and recollection.

#### ***4.3 Implications and Future Directions***

This current narrative manuscript, if developed into a shorter format for a peer-reviewed journal article, could encourage greater transparency about reporting guideline consensus meetings, as well as the use of anonymous, electronic voting technology in the future for such meetings. The votes and discussions from the consensus meeting will be used to re-word and re-arrange the final CONSORT-SPI checklist, as well as develop the content of the E&E document. Some standard CONSORT 2010 items may also be reworded to better fit concepts in the area to social and psychological intervention research. The name of the guideline may also be changed to better engage researchers who work on environmental, system-level, and structural interventions. Participants from the consensus meeting will help to finalise the CONSORT-SPI checklist, the CONSORT-SPI Extension Statement, and the CONSORT-SPI E&E. Interested Delphi participants and other stakeholders may also provide comments on drafts prior to publication. A first draft of the E&E template will comprise the next chapter of this thesis.

## References

1. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*. 2000;32(4):1008-1015.
2. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998;2(3).
3. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311(376).
4. Moher D, Weeks L, Ocampo M, et al. Describing reporting guidelines for health research: A systematic review. *Journal of Clinical Epidemiology*. 2011;64(7):718-742.
5. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
6. Mayo-Wilson E, Grant S, Hopewell S, Macdonald G, Moher D, Montgomery P. Developing a reporting guideline for social and psychological intervention trials. *Trials*. 2013;14:242.
7. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.
8. Montgomery P, Grant S, Hopewell S, et al. Protocol for CONSORT-SPI: An Extension for Social and Psychological Interventions. *Implementation Science*. 2013;8:99.
9. Moher D, Hopewell S, Schultz KF, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.

10. Hoffmann TC, Glasziou P, Boutron I, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687.
11. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *The Lancet*. 2008;371:281-283.
12. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
13. Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT statement by high impact factor medical journals: A survey of journal editors and journal 'Instructions to Authors'. *Trials*. 2008;9:20.
14. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698-702.
15. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ*. 2008 337:a2390.
16. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*. 2011;8(1):e1000393.
17. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT Group. Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: Explanation and Elaboration. *Annals of Internal Medicine*. 2008;148:295-309.

18. Kunz R, Fretheim A, Cluzeau F, et al. Guideline group composition and group processes: Article 3 in integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proceedings of the American Thoracic Society*. 2012;9:229-233.
19. Riessman CK. Narrative analysis. In: Huberman AM, Miles MB, eds. *The qualitative researcher's companion*. Thousand Oaks, CA: SAGE Publications, Inc.; 2002:216-271.
20. Mishler EG. *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press; 1986.
21. Corbin J, Strauss AC. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London, UK: Sage Publications; 2009.
22. *NVivo qualitative data analysis software* [computer program]. Version 10. 2012.
23. Reed J, Payton VR. Focus groups: Issues of analysis and interpretation. *Journal of Advanced Nursing*. 1997;26(4):765-771.
24. Behar R. *Translated woman: Crossing the border with Esperanza's story*. Boston: Beacon; 1993.
25. Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR Network's Survey of Guideline Authors. *PLoS Medicine*. 2008;5:869-874.
26. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews*. 2012;11:MR000030.

27. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: Defining standard protocol items for clinical trials. *Annals of Internal Medicine*. 2013;158(3):200-207.
28. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*. Oct 2009;62(10):e1-34.
29. APA JARS Group. Reporting standards for research in psychology: Why do we need them? what might they be? *American Psychologist*. 2008;63:839-851.
30. Glasgow RE. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*. 1999;89(9):1322-1327.
31. Shamseer L, Sampson M, Bukutu C, et al. CONSORT extension for N-of-1 trials (CENT) guidelines. *BMC Complementary and Alternative Medicine*. 2012;12(Suppl 1):P140.
32. Medical Research Council. *Developing and evaluating complex interventions: New guidance*. London: Medical Research Council; 2008.
33. Henggeler SW, Schoenwald SK, Rowland MD, Cunningham PB. *Serious emotional disturbances in children and adolescents: Multisystemic therapy*. New York: Guilford Press; 2002.
34. Booth A, Clarke M, Ghera D, Moher D, Petticrew M, Stewart L. Establishing a minimum dataset for prospective registration of systematic reviews: An international consultation. *PLoS One*. 2011;6(11):e27319.
35. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: Extension to cluster randomised trials. *BMJ*. 2012;345:e5661.

**Chapter 6:**  
**CONSORT-SPI Explanation and Elaboration:**  
**Template for the Checklist Users Manual**

**Abstract**

**Background:** Previous chapters have established that the quality of reporting randomised controlled trials (RCTs) of social and psychological interventions is sub-optimal. A Delphi process and subsequent consensus meeting were held to decide the content of the CONSORT-SPI checklist. Experts in the development and dissemination of reporting guidelines recommend that, to facilitate the proper use of reporting guidelines, an Explanation and Elaboration (E&E) document should be produced that provides examples of reporting according to checklist items and also explicates the rationale for each item.

**Methods:** To develop a draft template for the CONSORT-SPI E&E, examples of reporting according to the checklist were found for each item from the database of social and psychological intervention RCTs from the systematic review in Chapter 3 of this thesis, as well as other prominent trials in this area. The rationale for each reporting item is based on feedback from the CONSORT-SPI Delphi process and consensus meeting on what should be included in the E&E, as well as explanatory guidance from previous reporting guidelines in the social and behavioural sciences.

**Results:** The draft Explanation and Elaboration (E&E) template provided in this chapter builds on previous phases of this project to explicate the principles underlying checklist items and offers examples of adhering to these items.

**Conclusions:** The CONSORT-SPI checklist is intended to facilitate the writing, critical appraisal, and interpretation of social and psychological intervention RCTs. The E&E document will serve as a “users manual” for understanding the rationale of the CONSORT-SPI checklist and applying it appropriately when writing or appraising reports of social and psychological intervention RCTs. This template will be used by the various writing groups of consensus meeting participants developing E&E’s that are tailored to specific disciplines. Once refined via feedback from the Project Executive and these writing groups, the E&E should be a useful resource to help improve the quality of reporting social and psychological RCTs. In addition, by referring readers to empirical and theoretical studies underpinning these items in the final versions, the E&E may help to improve the design and execution of these trials as well.

## 1. Introduction

As described in the previous chapter, a three-day consensus development conference was held following a Delphi process in order to vote on items for the CONSORT-SPI checklist. Meeting participants voted on including 14 additional items for this new checklist. An increasingly recommended best practice when developing a new reporting guideline like CONSORT-SPI is to provide an Explanation and Elaboration (E&E) document along with the checklist. If guideline development groups only disseminate a checklist without guidance, there is concern that potential users may interpret such terse presentation as evidence of an insufficiently supported guideline that is dogmatically making prescriptions.<sup>1</sup> Experts in the development and dissemination of reporting guidelines therefore argue that publishing longer E&E documents alongside main reporting guideline statements is crucial to both instruct as well as persuade potential users of the importance of adhering to guideline recommendations.<sup>1</sup>

An E&E bolsters the support for all items in a checklist by providing (a) examples of good reporting from published papers, (b) relevant scientific background and rationale for including the requested information in published articles, (c) any empirical evidence of bias associated with the conduct or reporting of a study relevant to a checklist item, and (d) the extent of inadequate reporting of information related to checklist items.<sup>1</sup> To date, most reporting guidelines have not included E&E documents alongside (the shorter) main statements,<sup>2</sup> although the practice of writing a complementary E&E paper is growing.<sup>3</sup>

From the outset, the CONSORT-SPI Project Executive has intended to develop an E&E document, and to publish this document alongside the CONSORT-SPI checklist. Moreover, participants at the consensus meeting also indicated support for developing different versions of the E&E document that provide tailored examples and rationale to specific key disciplines targeted by the guideline. Several participants suggested that a

template should be developed prior to the drafting of these tailored E&E documents, in order to encapsulate the fundamental points consistently across all versions.

## **2. Methods**

The purpose of this chapter is to develop a draft template for the tailored versions of the CONSORT-SPI E&E document. The guidance provided below aims to capture insights from previous phases of the project in order to offer a template for the further development of CONSORT-SPI E&E documents that are tailored to specific disciplines.

### ***2.1 Procedure***

This E&E template follows the order of items in the draft CONSORT-SPI checklist, and its structure is based on the CONSORT 2010 E&E document.<sup>4</sup> To maximise the benefit of this document, readers are encouraged to use the E&E in conjunction with the CONSORT-SPI checklist.<sup>5</sup> In the current template, examples of reporting according to each item in the CONSORT-SPI checklist were found from the database of social and psychological intervention RCTs from the systematic reviews in Chapter 3 of this thesis. If no exemplar examples of adhering to a CONSORT-SPI checklist item were found from this database, the DPhil candidate located other social and psychological intervention RCTs in recent intervention literature. The explanation and elaboration for each item are based on previous phases of the project: E&E documents of previous reporting guidelines in the social and behavioural sciences (Chapter 3), feedback from the CONSORT-SPI Delphi process (Chapter 4), and decisions about E&E content at the consensus meeting (Chapter 5). Concepts requiring further explication are presented in tables and figures.<sup>4</sup> Discussion for some items also includes what is expected to appear in the final E&E.

Given the novelty of CONSORT guidance to many areas that CONSORT-SPI

targets, consensus meeting participants suggested that the CONSORT-SPI E&E explain the 14 additional items for the CONSORT-SPI checklist, as well as existing items in the CONSORT 2010 checklist. Consequently, all items on the CONSORT-SPI checklist—both those from CONSORT 2010 and the 14 items proposed during the consensus meeting—are discussed in this E&E template with social and psychological interventions in mind.

## ***2.2 Preambles to the E&E Requested by Consensus Meeting Participants***

Consensus meeting participants also noted several issues that should have a brief introduction in the E&E. These include the remit of the CONSORT-SPI guidance, using other CONSORT guidelines, how to use the checklist depending on the nature of the intervention techniques and maturity of previous evidence on them, the meaning of “participants” in the CONSORT-SPI checklist, reporting intervention implementation and process evaluations, and linking multiple documents on the same trial.

### ***2.2.1 Remit of CONSORT-SPI***

As stated earlier, social and psychological interventions are defined in this thesis as those interventions working through mechanisms that consist of mental processes and social phenomena. Such a definition corresponds to a variety of disciplines, theories, techniques, mechanisms, contexts, and outcomes potentially under the remit of CONSORT-SPI guidance. An indicative (though not comprehensive) list of these has been generated based on entries in the Campbell Collaboration library to give an idea of which intervention researchers may benefit from use of CONSORT-SPI guidance (see Table 1).

Consensus meeting participants also indicated the desire for a glossary to accompany CONSORT-SPI guidance. Participants flagged several key terms at the consensus meeting

**Table 1. Social and Psychological Interventions from the Campbell Library<sup>6</sup>**

<b>Interventions</b>	<b>Theories</b>	<b>Techniques</b>	<b>Settings</b>	<b>Mechanisms</b>	<b>Outcomes</b>	<b>Disciplines</b>
Active Labour Market Programmes	Equilibrium Unemployment Theory	Job search assistance	Social services departments	Job search efficiency	Exit from unemployment	Economics
Brief Motivational Interventions	Transtheoretical Model of Change	Motivational interviewing	Online	Motivation, knowledge, skills	Alcohol and drug use	Psychology
Cash-Based Approaches	Paternalism Theory	Cash-transfers	Humanitarian crises in LMICs	Increased purchasing power	Household income	Public Health
Gang Resistance Programmes	Self-Control Theory	Cognitive-behavioural training, education in conflict resolution	High schools	Social skills	Gang membership	Criminology
Mentoring Interventions	Social Learning Theory	Interactive relationship between mentor and mentee	Community organisations	Self-efficacy, societal bonds	Juvenile delinquency	Criminology
Montessori Education	Constructivism	Student-chosen school activities	Elementary schools	Experiential knowledge	Academic achievement	Education
"No Excuses" Charter Schools	Laissez-Faire Education	Rigorous, strict structure	Inner-city, urban areas	Knowledge and character	Literacy and numeracy	Education
Preventive Home Visits	Health Promotion Model	Multidimensional geriatric assessment	Homes in the community	Psychosocial and environmental risk factors	Institutionalisation	Public Health
Psychoanalytic psychotherapy	Psychoanalytic Theory	Talk therapy	Therapist offices	Insight on unconscious difficulties	Post traumatic stress disorder (PTSD)	Psychology
Skills-Based Intimate Partner Violence Prevention	Feminist Theories of Gender Inequality	Modelling positive relationship skills	Universities	Communication skills, self-esteem	Dating violence	Social Work

to include in such a glossary. For example, many terms related to any theories underpinning how the intervention is hypothesised to work, such as “mechanism of action”, “programme theory”, “logic model”, “conceptual framework”, and “formal theory”. Several other terms relate to the implementation of interventions, for example, “fidelity”, “integrity”, “tailoring”, “adaptation”, and “modification”. Further terms will likely be identified as the Project Executive and consensus meeting participants revise the E&E document moving forward. All terms will be collated into one list containing clear, concise definitions, for inclusion in the published version of the CONSORT-SPI E&E.

### *2.2.2 Other CONSORT Extensions*

Several other CONSORT extensions may be relevant and beneficial to social and psychological intervention researchers for use in conjunction with CONSORT-SPI. Firstly, as noted below in Item 1b, the CONSORT for Abstracts Extension<sup>7</sup> will be useful for the title and abstracts of social and psychological intervention trial publications and conference submissions (Box 1 under Item 1a below). For those employing a cluster trial design, the CONSORT Extension for Cluster Randomised Trials<sup>8</sup> should be used as well (see Table 3 under Item 3a below). The cluster extension may also prove useful to those reporting trials with hierarchical data structures even if the trial is not cluster-randomised, such as individually randomised trials of students nested within classrooms nested within schools. Given the prevalence of cluster trials within the social and psychological intervention literature, CONSORT 2010 items with extended cluster extension items have been flagged in the CONSORT-SPI checklist (see Table 2 below). Moreover, an appendix of the CONSORT-SPI extension may include a modified checklist that incorporates elements from the cluster extension. For those employing medical interventions as a comparator to or in conjunction with social and psychological interventions, the CONSORT Extension

for Non-Pharmacologic Treatments<sup>9,10</sup> may prove useful to consult.

### *2.2.3 Nature of the Intervention Techniques and the Trial*

The detail desired in the E&E for each of the CONSORT-SPI checklist items may differ depending on the specific nature of the intervention techniques being evaluated. For example, discrete intervention techniques targeted at individuals may differ in how they adhere to items compared to large-scale social policy interventions. How to adhere to these items will also depend on the phase of the trial and its key research questions. When relevant, the template below indicates how reported information may differ for interventions at the piloting and feasibility, efficacy, effectiveness, or implementation stages, as appropriate. This model should be followed in the final E&E documents.

Consensus meeting participants noted that the phased approach to developing and evaluating social and psychological interventions is not as judiciously followed in social and psychological intervention research as in pharmaceutical areas, where strict regulation enforces a phased development. As such, many social and psychological interventions skip certain phases, so details that typically would not be provided for a trial at one phase may need to be reported considering the lack of corresponding earlier phases or planned later phases. For example, implementation details may be crucial to report in an efficacy trial—i.e., a tightly controlled trial under ideal conditions where fidelity to protocol is rigorously pursued<sup>11</sup>—when the results may immediately lead to policy decisions or wide-scale replication. Authors will need to use their judgment to adhere to items according to the nature of the intervention and purpose of their trial.

### *2.2.4 Meaning of “Participants” in the CONSORT-SPI Checklist*

In accordance with the headers in the CONSORT 2010 guidance,<sup>4</sup> the CONSORT-

SPI checklist and E&E will refer to the unit targeted by the intervention as “participants”.

While this term tends to connote individual people, not all social and psychological interventions target individuals. Instead, many target groups of individuals such as families or schools, or even involve place-based techniques that intervene on geographic units and examine area-level effects. As such, “participants” throughout this guidance is meant to stand for “participating units”, which involve the unit being targeted by the intervention.<sup>12</sup>

To facilitate this understanding of “participants”, a few adjustments have been preliminarily made to the CONSORT-SPI checklist as detailed in the section “Changes to CONSORT 2010 Checklist for CONSORT-SPI” below.

### *2.2.5 Intervention Implementation and Process Evaluations*

Stakeholders in social and psychological intervention research often want to know how, why, and under what circumstances interventions work, rather than just whether they work. Process evaluations are increasingly embedded within social and psychological RCTs, as resources allow, in order to understand the functioning of an intervention by examining its implementation, the targeted causal mechanisms, and relevant contextual factors.<sup>13</sup> Several explicit checklist items related to process evaluations were considered during the CONSORT-SPI Delphi process and consensus meeting. However, feedback from these consensus methods indicated that not all social and psychological intervention trials have the resources or ability to conduct process evaluations, nor are process evaluations always needed or appropriate. As such, explicit items about process evaluations are not in the checklist, though related topics will be discussed in the E&E. The checklist has addressed what needs to be reported in the main publication for all social and psychological intervention trials. The E&E text will then go into further detail about what might be expected in sibling documents, such as linked process evaluation reports.

However, where possible, critically important items should be addressed within or concurrent with the main trial paper, as intended additional manuscripts often do not get published as planned, if at all. Such information can be provided either in the text, as some kind of web appendix, or via a reference to another publication.

#### *2.2.6 Record Linkage and Sibling Documents*

Related to the issue of process evaluations is the linking of all documents reporting on the same trial. Given the vast undertaking that many social and psychological intervention trials are, many have proposed that it is impossible to provide all of the desired information about a trial in a single publication, either due to journal word limits or the greater time it takes to conduct some analyses than others. For a single trial, then, several publications could potentially exist, such as a protocol paper, an intervention design paper, a primary outcomes paper, a process evaluation paper, and papers reporting secondary analyses of outcomes. Moreover, each report may have a different authorship order or analyse different aspects of the dataset, making it difficult for readers to identify whether the same trial is discussed across documents. The CONSORT-SPI E&E guidance should aim to indicate which information should be reported in the main outcomes paper, and then discuss other information that may be in an appendix, web resource, or reported in other sources given word counts and common publication practices.

Though noted in Item 23 below, it is further emphasised in this section that trialists link all records about a trial. It is primarily recommended that authors register trials in advance, and update trial registration entries with citations to all manuscripts and reports of a trial. Citing all previous papers on a trial consecutively as manuscripts emerge can lead to issues of self-citation, too many citations (if several previous manuscripts exist), and inability to discover later papers when an interested reader has accessed earlier ones. In

contrast, trial registration numbers provide a unique identifier for the study that can be cited across all documents.

### ***2.3 Changes to the CONSORT-SPI Checklist After the Consensus Meeting***

For this thesis chapter, the DPhil candidate has revised the draft CONSORT-SPI checklist at the end of the consensus meeting based on the analysis of meeting discussions presented in the last chapter. A full checklist of CONSORT-SPI items mapped onto the CONSORT 2010 Statement can be found in Table 2 below.

In response to many suggestions at the consensus meeting, some of the additional items proposed for the CONSORT-SPI checklist have been altered. For example, the item “Why and how the intervention is hypothesised to work” has been changed to “If pre-specified, how the intervention was hypothesised to work” and moved from Item 2a to Item 2b. Consensus meeting participants indicated that “Why and” was a confusing wording for the item, and that the item seems to be better placed as an extension of CONSORT 2010 Item 2b on the study hypothesis. In addition, the CONSORT-SPI Item 5 has been changed to the “Extent to which interventions were delivered and taken up as planned, including what they actually involved”, adding the clause “including what they actually involved” to respond to consensus meeting participants’ desire to emphasise “what” the intervention actually entailed.

In addition to these changes, certain aspects of the CONSORT 2010 checklist format and items have also been *preliminary* amended by the DPhil candidate to reflect concerns by consensus meeting participants about the CONSORT 2010 checklist’s suitability to social and psychological interventions. These reflect the type of changes indicated in “Box 2. Noteworthy specific changes in CONSORT 2010 Statement” in the CONSORT 2010 update.<sup>14</sup> It cannot be underemphasised that these alterations are not

**Table 2. Draft CONSORT-SPI Checklist (Stage: Post-Consensus Meeting)**

Section	Item #	Standard CONSORT Description	Proposed CONSORT-SPI Items
<b>Title and abstract</b>			
	1a	Identification as a randomised trial in the title <sup>§</sup>	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts) <sup>§</sup>	
<b>Introduction</b>			
Background and objectives	2a	Scientific background and explanation of rationale <sup>§</sup>	
	2b	Specific objectives or hypotheses <sup>§</sup>	If pre-specified, how the intervention was hypothesised to work
<b>Methods</b>			
Trial Design	3a	Description of trial design (such as parallel, factorial) including allocation ratio <sup>§</sup>	If the unit of random assignment is not the individual, please refer to CONSORT for Cluster Randomised Trials
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants <sup>§</sup>	When applicable, eligibility criteria for settings and those delivering the interventions
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered <sup>§</sup>	Extent to which interventions were delivered and taken up as planned, including what they actually involved
			*Where other informational materials about delivering the intervention can be accessed
			When applicable, how intervention providers were assigned to each group

Section	Item #	Standard CONSORT Description	Proposed CONSORT-SPI Items
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed <sup>§</sup>	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample Size	7a	How sample size was determined <sup>§</sup>	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
<i>Randomisation:</i>			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size) <sup>§</sup>	
Allocation Concealment Mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned <sup>§</sup>	
Implementation	10	Where applicable, who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions <sup>§</sup>	
Awareness of Assignment	11a	Who was aware after assignment to interventions (for example, participants, providers, those assessing outcomes), and how any masking was done	
	11b	If relevant, description of the similarity of interventions	
Analytical Methods	12a	Statistical methods used to compare groups for primary and secondary outcomes <sup>§</sup>	How missing data were handled (e.g., complete case analysis, simple imputation, multiple imputation), with details of any imputation method
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	

Section	Item #	Standard CONSORT Description	Proposed CONSORT-SPI Items
<b>Results</b>			
Participant Flow (a diagram is strongly recommended)	13a	For each group, the numbers randomly assigned, received intended treatment, and analysed for the primary outcome <sup>§</sup>	Where possible, the number approached, screened, and eligible prior to random assignment, with reasons for dropout
	13b	For each group, losses and exclusions after randomization, together with reasons <sup>§</sup>	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline Data	15	A table showing baseline characteristics for each group <sup>§</sup>	Including socioeconomic variables where applicable
Numbers Analysed	16	For each group, number included in each analysis and whether the analysis was by original assigned groups <sup>§</sup>	
Outcomes and Estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval) <sup>§</sup>	*Indicate availability of trial data
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary Analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
<b>Discussion</b>			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings <sup>§</sup>	

Section	Item #	Standard CONSORT Description	Proposed CONSORT-SPI Items
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
<b>Important Information</b>			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Declaration of Interests	25	Sources of funding and other support, role of funders	Declaration of any other potential interests
*Stakeholder Involvement	New Item		*Any involvement of the intervention developer in the design, conduct, analysis, and reporting of the trial *Other stakeholder involvement in trial design, conduct, and/or analyses *Incentives offered as part of the trial

\*Indicates item might move to another section

§ Indicates that an extension item for cluster trials exists

approved changes by the CONSORT Group or the CONSORT-SPI Project Executive, but rather represent the DPhil candidate's attempts to be responsive to participant comments from the consensus meeting for the purposes of this thesis. The final version of both the checklist and the E&E will undoubtedly involve many revisions of the template presented below and will ultimately need approval from the CONSORT Group.

Firstly, "Blinding" has been changed in Item 11 to "Awareness of assignment" and "masking" in the section heading and item wording, respectively. This change addresses concerns about the use of the term "blinding" as well as the need to emphasise the issue of awareness of assignment by providers and participants in social and psychological intervention trials. Secondly, the Item 12 section heading "Statistical methods" has been changed to "Analytical methods", given that some methods may be qualitative in these trials. For the wording in Items 13a and 16, "number of participants" has been changed to "number" in line with Section 2.2.4 above on the meaning of "participants". The parenthetical "(denominator)" has also been removed from Item 16, as this term seemed to accentuate the use of dichotomous outcomes, whereas continuous outcomes are extremely prevalent in this trial literature. In the wording of Item 15, "clinical and demographic" have been removed to avoid potentially perceived medical language, as well as emphasise the need to report all baseline characteristics collected, irrespective of their nature. The section "Other Information" has been changed to "Important Information" due to consensus meeting participants' concerns that "Other" makes the requested information seem to be of secondary importance to other sections. In the wording of Item 25, the phrase "such as supply of drugs" has been removed, as drug trials are not in the purview of this extension by definition. Lastly, a new sub-section in "Important Information" called "Stakeholder Involvement" has been added, as consensus meeting participants thought such a sub-section would best fit the three sub-items currently allocated to it.

### 3. Results: CONSORT-SPI Explanation and Elaboration

#### 3.1 Title and Abstract

##### *Item 1a. Identification as a randomised trial in the title*

*Example:* “Efficacy and effectiveness of individual family intervention on social and clinical functioning and family burden in severe schizophrenia: A 2-year randomized controlled study”<sup>15</sup>

*Explanation:* The title signals to the potential reader whether the study might be relevant to their research needs. As the unique characteristic of RCTs, identifying a trial as “randomised/randomized” in the title signifies the centrality of the methodology to the document content. Full document searches are not always contained in systematic review search strings, but may instead focus on the title and abstract to make the number of documents retrieved more manageable. As such, systematic reviewers and others scanning the research literature can more easily and quickly identify randomised trials if the term is in the title.<sup>16</sup> Placing “randomised” in the title may also help those organising research databases to properly index the trial.<sup>16</sup> Though authors in some disciplines use the term “experiment” alone without the term “randomised” because “experiment” tends to imply randomisation,<sup>17</sup> use of a term including “random” in the title is suggested because some interested readers may not be aware of this implication. Lack of standardised use of this term in the title (and abstract) can impede the retrieval of the relevant social and psychological intervention evidence base.<sup>18</sup>

Preferably, if character limits allow, authors should also consider providing other details about an RCT in the title, including the intervention being evaluated and the problem being targeted by the intervention. Such information facilitates searching and screening of the literature by researcher users, particularly systematic reviewers, as it indicates the key variables and theoretical issues under investigation in the trial.<sup>19</sup> To stay within journal character limits, authors may use less precise terms for the intervention and

population than are used in the abstract and full manuscript, though these terms should be in accord with common terminology in the research literature. Authors are recommended against using uninformative catchy phrases in titles, as these can take up space better filled by key details that are more likely to be identified through electronic searches.<sup>16,19</sup>

*Item 1b. Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)*

*Example:* “Background: Recent guidelines for the treatment of irritable bowel syndrome (IBS) emphasize the need for research to facilitate home-based self-management for these patients in primary care. The aim of the current study was to test the efficacy of a manualized cognitive behavioural therapy (CBT)-based self-management programme for IBS in a pilot randomized controlled trial (RCT).

Method: Sixty-four primary-care patients meeting Rome criteria for IBS were randomized into either self-management plus treatment as usual (TAU) (n=31) or a TAU control condition (n=33). The self-management condition included a structured 7-week manualized programme that was self-administered in conjunction with a 1-hour face-to-face therapy session and two 1-hour telephone sessions. The primary outcome measures were the Subject’s Global Assessment (SGA) of Relief and the Irritable Bowel Syndrome Severity Scoring System (IBS-SSS) assessed at baseline, end of treatment (2 months), and 3 and 6 months post-treatment.

Results: Analysis was by intention-to-treat. Twenty-three (76.7 %) of the self-management group rated themselves as experiencing symptom relief across all three time periods compared to seven (21.2 %) of the TAU controls [odds ratio (OR) 12.2, 95% confidence interval (CI) 3.72–40.1]. At 8 months, 25 (83%) of the self-management group showed a clinically significant change on the IBS-SSS compared to 16 (49 %) of the control group (OR 5.3, 95 % CI 1.64–17.26).

Conclusions: This study provides preliminary evidence that CBT-based self-management in the form of a structured manual and minimal therapist contact is an effective and acceptable form of treatment for primary-care IBS patients.”<sup>20</sup>

*Explanation:* Abstracts are probably the most widely read section of a manuscript, and many readers use the abstract to decide whether to read the full manuscript.<sup>19</sup> As with the title, the abstract informs the reader of whether the full article is likely to be of interest, and it includes key information for properly indexing articles so that they are retrieved in electronic database searches.<sup>16</sup> Given greater character limits, authors should provide more

information and use more precise terms for key details in the abstract than in the title. The CONSORT Extension for Abstracts<sup>7</sup> and its E&E document<sup>21</sup> provide detailed advice for writing accurate, concise, and coherent abstracts of RCTs for both journal and conference submissions (see Box 1).

The abstract should include information about study methods, such as random assignment, sample size, and outcome measures. Authors should also note the problem or issue under investigation, sample characteristics, and all interventions in experimental and comparison groups. General aspects of the setting, such as geographic region, are important for potentially interested readers conducting a systematic review, planning a trial, or making decisions about policy and practice, particularly for trials intended to have an international audience. Authors should note the primary findings of the trial, including effect size estimates with confidence intervals. The abstract should end with a brief conclusion and statement of implications. Structured abstracts with headings are recommended to make the information easier to scan.

### **3.2 Introduction**

#### *Item 2a. Scientific background and explanation of rationale*

*Example:* “The disturbingly high rates of sexual assaults are clearly documented. National data suggest that 15% to 25% of women will be the victim of an attempted or completed rape during their lifetime. Research suggests college women are at greater risk for sexual victimization than women in the general population. The mental health consequences of sexual assault are serious. Women who are victims of sexual violence have higher and more severe rates of posttraumatic stress disorder (PTSD) than survivors of accidents and natural disasters. In addition to PTSD, there are many other insidious effects of sexual violence, which include psychological distress, physical distress, interpersonal problems, and increased risk for sexual revictimization”<sup>22</sup>

*Explanation:* The introduction section of a trial report situates an RCT in existing literature. A clear and structured introduction section concisely details the rationale underpinning the study and persuades the reader as to how the study will importantly

contribute to what is already known.<sup>18</sup> A brief, up-to-date synopsis of relevant intervention literature, preferably referencing any systematic reviews in the area, should be provided for the reader to understand how this trial relates to and builds on previous work.<sup>19</sup> With the context and setting of the trial in mind, this summary should aim to include information about the targeted outcome, the risk and protective factors related to the outcome, and previous information about interventions to address this outcome—in particular focusing on the intervention under investigation.<sup>23</sup>

The problem or issue necessitating the work should be stated and its significance explained, such as its prevalence, incidence, severity, or course without intervention, to indicate the need for intervention.<sup>24</sup> Theories about the nature of the targeted issue or problem would also be helpful to provide. Authors may also address limitations of previous research as well as existing interventions.<sup>18</sup> Social and psychological interventions often have overlapping components, so clearly demarcating the intervention under study from other alternatives—most importantly the comparator in the study—is needed to clarify specifically what techniques are being evaluated.

Current knowledge on the intervention will vary by maturity, though authors should comprehensively summarise the totality of knowledge about the intervention rather than selectively presenting studies favouring a particular point of view. If the experimental intervention is new or a novel adaptation of an established intervention, authors should report why other interventions are insufficient and how this intervention may differ, particularly if any other interventions have evidence of effectiveness or serve as comparators in the trial. Authors should also report the development and any previous piloting of the new intervention if done, with reference to any relevant manuscripts. If a more established intervention is being evaluated, authors should summarise the current state of the evidence for the intervention (e.g., referencing previous evaluations or

systematic reviews of its effectiveness) and the need for this study (e.g., adapt it to a new setting or population, resolve inconclusive evidence). Consistent with this is an explanation for the choice of comparator in the trial. For all of the above, references to systematic reviews when existent are preferred to primary studies.

*Item 2b. Specific objectives or hypotheses*

*Example:* “It was proposed that an individual intervention that used cognitive-behavioral and motivational interviewing approaches to enhance coping and self-regulation among HIV-positive MSM (men who have sex with men) would be effective in fostering their sexual safety.... We thus hypothesize that sexual safety among HIV-positive men would be facilitated by self-efficacy and skills for enhancing social support and coping with HIV, modulating negative affect, enhancing HIV disclosure, and enhancing information and motivation specifically around sexuality.... We [also] hypothesized that overall unprotected anal intercourse [UAI] would lessen only moderately, whereas transmission risk—UAI that may transmit HIV to uninfected partners—would show significant intervention effects. We anticipated that intervention effects would be strongest at the 6-month, follow-up point and would remain significant at 12 months.”<sup>25</sup>

*Explanation:* Objectives specify the aims of a study, whereas hypotheses propose specific research questions with expectations, such as the magnitude and direction of effects. These may differ depending on the nature of a trial in piloting a new intervention, enhancing understanding about its efficacy and hypothesised mechanisms, or determining its effectiveness as a potential practice to be scaled up.<sup>24</sup> Objectives and hypotheses should follow on the empirical evidence and theory provided for the rationale of the trial, and the research design should clearly be able to address them. Rather than making general statements, authors should specify objectives or hypotheses in a clear and unambiguous way in terms of the sample or target population, the nature of the interventions in the context of the trial, and the assessed outcomes within the follow-up time-frame in the trial.<sup>18</sup> Authors should note which objectives or hypotheses are primary (i.e., the main focus of the intervention trial) and which are secondary to provide context for interpreting the relative importance of findings across outcomes in a trial. Those findings deemed

primary and specified in advance have more credibility than those deemed secondary or tested *post hoc*.<sup>19</sup> Because some social and psychological interventions have various units of intervention and outcome, such as individuals, clusters, and places, authors should specify to whom or what particular objectives or hypotheses apply. Authors should also indicate whether hypotheses were pre-specified.

*Extended CONSORT-SPI Item 2b. If pre-specified, how the intervention was hypothesised to work*

*Example:* “Feedback provides students with information about their learning and achievement and gives learners the opportunity to adjust and develop their cognitive strategies and to rectify misconceptions while progressing through the training. Feedback has been argued to play an important role in learning and to influence performance in different ways depending on how it is provided. Especially during initial practice, feedback should be provided during each step of the problem-solving procedure. This allows learners to immediately verify the correctness of a solution step while the corresponding step is still in working memory, and it enables learners to focus their attention on the solution steps.”<sup>26</sup>

*Explanation:* Authors should clarify the conceptual underpinnings of the intervention being evaluated in the trial.<sup>19</sup> Various terms exist for identifying how an intervention is hypothesised to work, such as its theory of change, change or causal processes, programme theory, causal pathways, and mechanisms of action.<sup>27</sup> Being clear about what the intervention is thought to impact and how this leads to changes in outcomes allows the reader to appraise the strength of the theoretical basis of the intervention as well as the appropriateness of the trial’s methods (e.g., choice of outcome measures). Basing the design of a trial on an explicit theoretical framework can help an RCT move beyond detecting an effect to actually explaining the effect of the intervention and generating cumulative scientific knowledge.<sup>12</sup> However, if the intervention is not based on theory, authors should not report programme theories *post hoc*.<sup>12</sup> Rather, authors should simply state if no mechanisms were explicitly hypothesised prior to the trial.

If a hypothesis was pre-specified about how the intervention works, authors should be clear about the choice of outcomes, mechanisms, and active components of the intervention (e.g., techniques, activities employed); how intervention components impact the targeted mechanisms; how impacting these mechanisms leads to a change in outcomes; and the role of contextual features in this process. A full consideration of the hypothesised barriers to intervention effectiveness as well as potential adverse events or unintended consequences may also be helpful to discuss.<sup>18</sup> Any substantive or “formal” theories (e.g., the Theory of Planned Behaviour) used to design the intervention and underpin the hypothesised functioning of the intervention should be mentioned.<sup>12</sup> For particularly complex interventions, logic models or conceptual frameworks may be helpful in depicting how these interventions are intended to work. Such graphical depictions are especially useful for those interventions that work through various mechanisms at individual and group levels, as well as those that have long causal chains (see Figure 1).

### **3.3 Methods: Trial Design**

*Item 3a. Description of trial design (such as parallel, factorial) including allocation ratio*

*Example:* “We employed a 2 (intervention: CBT (cognitive-behavioural therapy) vs. GHE (general health education)) x 3 (time: end of counseling at 2 weeks, follow-ups at 3 and 6 months) mixed factorial design. Participants ... were randomly assigned to receive either CBT or GHE at a 1:1 ratio ( $n = 77$  per intervention group).”<sup>28</sup>

*Explanation:* Specific and unambiguous details on the design of the trial clarify to the reader the suitability of the study to address study objectives or hypotheses.<sup>19,23</sup> Authors should make clear both the level of random assignment as well as the level of analysis. Complex designs, such as stepped-wedge or Zelen designs, should be described if used. If relevant, official CONSORT extensions exist for non-inferiority and equivalence trials,<sup>29</sup> as well as pragmatic trials,<sup>30</sup> with a new extension for N-of-1 trials currently under

development.<sup>32</sup> The design decisions (e.g., the nature of the comparator) should indicate the phase of the trial commensurate with the maturity of the intervention being tested.

Authors should also make clear the way in which any theories underlying the intervention and how it works have been applied to the design and analysis of the trial.

*Extended CONSORT-SPI Item 3a. If the unit of random assignment is not the individual, please refer to CONSORT for Cluster Randomised Trials*

*Example:* “We conducted a paired randomized school-based trial in a sample of fifth graders from public schools in the municipality of Duque de Caxias, Rio de Janeiro, Brazil. The study was conducted and reported according to the CONSORT guideline for cluster-randomized trials (Campbell et al., 2012). The protocol for this trial and supporting CONSORT checklist are available as supporting information.”<sup>33</sup>

*Explanation:* Authors should provide various additional details throughout the report if the trial was cluster randomised. For further guidance, authors should consult the CONSORT Extension to Cluster Randomised Trials.<sup>8</sup> See Table 3 below for a list of the extended items in that checklist.

*Item 3b. Important changes to methods after trial commencement (such as eligibility criteria), with reasons*

*Example:* “Originally, each teacher had three seventh-grade classes participating in the study. Each class was randomly assigned to receive one of the treatments so that each teacher taught all three treatments. The purpose of this restricted random assignment was to reduce the potential confounding effect of instructional methods. However, after the school year started, Teacher C lost one of her seventh-grade classes (the third treatment) because of a change of assignment to teach eighth-grade science.”<sup>34</sup>

*Explanation:* Trials are rarely executed exactly as designed. Reporting changes to or deviations from the trial protocol after the trial has begun is important to interpreting the results of a trial. With reference to a trial protocol, authors should note any important changes after the trial has started that could influence the quality or applicability of results. For those who did not develop a protocol or register their trial in advance, see the discussion for Items 23 and 24 below for guidance.

**Table 3. Extended Items from CONSORT for Cluster Randomised Trials<sup>8</sup>**

Item	Section	Cluster Extension Item
1a	Title	Identification as a cluster randomised trial in the title
2a	Background	Rationale for using a cluster design
2b		Whether objectives pertain to the cluster level, the individual participant level, or both
3a	Trial Design	Definition of cluster and description of how the design features apply to the clusters
4a	Participants	Eligibility criteria for clusters
5	Interventions	Whether interventions pertain to the cluster level, the individual participant level, or both
6a	Outcomes	Whether outcome measures pertain to the cluster level, the individual participant level, or both
7a	Sample Size	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or $k$ ), and an indication of its uncertainty
8b	Sequence Generation	Details of stratification or matching if used
9	Allocation Concealment Mechanism	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level, or both
10a	Randomisation Implementation	Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions
10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)
10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both) and whether consent was sought before or after randomisation
12a	Statistical Methods	How clustering was taken into account
13a	Participant Flow	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome
13b	Participant Flow	For each group, losses and exclusions for both clusters and individual cluster members
15	Baseline Data	Baseline characteristics for the individual and cluster levels as applicable for each group
16	Numbers Analysed	For each group, number of clusters included in each analysis
17a	Outcomes and Estimation	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or $k$ ) for each primary outcome
21	Generalisability	Generalisability to clusters and/or individual participants (as relevant)

### 3.4 Methods: Participants

#### *Item 4a. Eligibility criteria for participants*

*Example:* “Families were referred to the program by schools (30%), community-based agencies (22%), health care clinics (21%), self (16%), or public social services (12%). Participants were screened according to the criteria listed above and recruited during the 4-year period (1997–2001). Of the 302 families screened, 216 met the eligibility criteria.... The targeted families resided in Baltimore’s Westside Empowerment Zone (i.e., federally designated as an area of extreme poverty, unemployment, and general economic distress) and had at least one child between the ages of 5 and 11. Eligibility for the FC program included (a) a concern by the referring person that at least 1 of 19 neglect subtypes (e.g., unsafe housing conditions, inadequate supervision, inadequate/delayed health care) was occurring at a low level but not at a level that Child Protective Services (CPS) would accept for investigation; (b) at least two additional risk factors for neglect related to the child (e.g., behavioral problem; physical, developmental, or learning disability; more than three children) or the caregiver/family (e.g., unemployment/ overemployment, mental health problem, drug or alcohol problem, domestic violence, homelessness); (c) no current CPS involvement; and (d) caregiver expressed willingness to participate in the FC program.”<sup>35</sup>

*Explanation:* Authors should provide all criteria—both inclusion and exclusion—that were used to determine the eligibility of participants in the trial, particularly if these relate to important demographic characteristics.<sup>19</sup> Eligibility criteria help readers understand the nature of the participants in a study and thus to whom the results of trials may or may not apply. Reported eligibility criteria should match a referenced and accessible trial protocol, and reasons for any post-hoc changes or violations should be explained. Explicit reporting of any exclusion criteria in addition to inclusion criteria is essential to know which populations that are related to the study samples were not included. Eligibility criteria for groups and places (e.g., families, schools, practice centres, geographic areas) should also be provided where these are the units being targeted by the intervention. Authors should also make clear how they operationalised screening processes according to the eligibility criteria, such as methods of diagnosing participants to be at-risk of a mental disorder.

As trials almost never use random sampling procedures, a description of the method of recruitment or sample is also important in order to better understand the nature of the

target population. The methods used to approach and screen participants reveal the structure and processes that led to them to be in the study. From this, readers can understand whether the sampling frame matches the target population(s) of the trial, and how recruitment practices relate to the sample that was ultimately included in the trial. These details guide the reader in making judgements about the restrictions on who might have entered the trial and thus the generalisability of findings, which are dependent on the recruitment scheme and characteristics of the actual sample. Authors should indicate the recruitment settings as well as the methods of recruitment and referral employed in the study, indicating which are reactive (e.g., participant self-selection) or proactive (e.g., approaching schools) in finding potential participants.<sup>18</sup> Differences in effect sizes may be found across trials assessing the same intervention with samples that were universally screened compared to those trials that selectively and opportunistically sampled participants who proactively were seeking to change. In addition, certain practices of approaching participants—such as coercion in trials conducted in criminal justice systems, or conveniently sampling college students in introductory psychology classes—may lead to restricted sub-samples of the targeted population.<sup>19,36</sup>

*Extended CONSORT-SPI Item 4a. When applicable, eligibility criteria for settings and those delivering the interventions*

*Example:* “The following study inclusion criteria were used to control for any further variability in classroom characteristics across program type: (1) teachers with a bachelor’s degree or an associate’s degree and working towards a bachelor’s; (2) programs with moderate to high quality as measured by the NC (North Carolina) star-rating system (3–5 stars out of 5 stars total), (3) use of the Creative Curriculum (a state-approved MAF (More at Four Pre-Kindergarten Program) curriculum and the predominant curriculum used by MAF classrooms), (4) classroom enrollment of at least four Latino ELL (English Language Learner) children, but not to exceed 85% of total enrollment, and (5) use of English as the primary language of instruction.”<sup>37</sup>

*Explanation:* In addition to the criteria used to recruit participants, trials may also have

eligibility criteria at different levels of the recruitment or sampling plan.<sup>12</sup> To the same degree as eligibility criteria for participants, trialists should indicate the criteria used to select those providing the intervention as well as the settings in which the intervention was delivered, when such criteria exist. This information helps the reader compare the context of the trial with the specific contexts of interest to them,<sup>18</sup> as estimated effects of the intervention can depend on the settings in which the intervention takes place and those delivering the intervention.<sup>38</sup>

*Item 4b. Settings and locations where the data were collected*

*Example:* “We were able to partner with the Charles A. Dana Center at the University of Texas, which has both a strong interest in increasing the number of students who progress in mathematics to advanced placement (AP) calculus and a history of providing teacher professional development at large scale throughout the state. Through its highly regarded and extensive professional development programs for mathematics teachers in Texas, the Dana Center had the ability and credibility to recruit teachers throughout the state, to help the project address concerns that potential teachers and administrators might have about participation in the project, and to facilitate workshops and provide workshop components, including a train-the-trainer model. In addition, the Dana Center had already been promoting an aligned sequence of instruction leading from middle school through AP calculus, and SimCalc naturally fit into this sequence. Third, Texas has an established, stable, and well-aligned system of standards and accountability. This enabled us to align the SimCalc curriculum with existing Texas curricula and standards and have confidence that the standards would not change midstream in our research and that conversations with teachers about curriculum would be consistent throughout the state. Fourth, Texas conducts a yearly census of teachers, schools, and districts. This allowed us to evaluate the properties of our sample relative to more general demographic information.”

*Explanation:* Readers need information about setting to understand the role of context in a trial, which helps in attempts to isolate and understand the effects of an intervention versus its comparator. It may be difficult to reliably tell which aspects of context, if any, are important to understand a social or psychological intervention. Authors should use their judgment when considering which aspects of context might be relevant to report.

Programme theory and other relevant evidence can be useful for deciding what to report and determining how an intervention functioned in the trial context. For example, it is

particularly important to discuss aspects of the context needed for interventions to work, or possible significant feedback mechanisms between the intervention and aspects of the external environment. Moreover, context may change over the course of intervention implementation, because of natural evolution or even feedback mechanisms between the intervention and context, for example.

As a foundation, concrete details of the setting may be plausible and helpful to report. When reporting such information about the setting, however, authors should be aware of confidentiality concerns, and not report information that can lead to people or organisations being identified when such identification would harm participants. Authors may note the geographic location of the trial (e.g., city) and day and time of trial activities (e.g., workday mornings in October). They can also note the key features of settings that are directly related to experiences of providers and participants, such as the physical space used to run an intervention. Information about organisations implementing interventions is also helpful, such as the educational philosophy of a school, or whether competing administrative priorities exist in a prison. At a wider level, relevant features of the external environment of the trial could be reported, such as community demographics or social care policies in place at the time of the trial. Such details about context are of interest particularly to those considering an intervention for policy and practice.

### ***3.5 Methods: Interventions***

*Item 5. The interventions for each group with sufficient details to allow replication, including how and when they were actually administered*

*Example:* “Participants received brief telephone calls for up to 18 months. These 5–10-min calls were offered weekly for the first 8 weeks, every other week for the next 44 weeks, and once per month for the final 6 months. Therefore, the total number of possible scheduled calls in the protocol was 36. Each call consisted of a structured 10-item assessment of current substance use status, other risk factors (e.g., craving, low self-efficacy, depression), and protective factors (e.g., attendance at self-help meetings, participation in other prorecovery social activities), which was referred to as a progress

assessment. A scoring algorithm produced a single summary score, with three levels of risk (e.g., low, medium, high), which was provided to the participant. A score of high risk could be generated by scores on two items (e.g., any substance use in the past week, very low confidence in being able to abstain from use in the coming week) or a combination of high scores on other risk factors and low scores on protective factors. A score of low risk required some participation in prerecovery behaviors, in addition to low scores on the risk items. In the TM condition, the calls did not include any formal counseling. However, when participants were relapsing or experiencing other crises, the counselors did make recommendations to reenter treatment or go to other facilities for help as part of the feedback they provided. In a few cases, counselors also suggested that a patient call back prior to his or her next scheduled contact. However, in general no stepped care was offered in this condition.”<sup>39</sup>

*Explanation:* As the independent variable in an RCT, a comprehensive summary of all interventions—experimental and comparator—should be provided in a trial report. It is critical to provide readers with sufficient detail to understand the content and delivery of all interventions provided in a trial.<sup>12</sup> However, as space is limited in many journals, full information about the design of the intervention can be provided in other materials (see extended CONSORT-SPI Item 5 below).<sup>12</sup>

Authors should consult the Template for Intervention Description and Replication (TIDieR) for summarising the interventions tested (see Table 4).<sup>40</sup> TIDieR acknowledges the complex nature of many interventions, including social and psychological interventions, and provides guidance on the essential aspects to report so that others can fully understand an intervention. These involve naming the intervention and then, as applicable, describing why it is supposed to work (CONSORT-SPI Item 2b above), what is actually delivered (materials and procedures), who provides the intervention, how they provide it, where they provide it, when they provide it, and an indication of how much is provided. Those using the TIDieR checklist to describe their interventions should use the most up-to-date version, checking to see whether TIDieR has been updated. For some complex interventions—such as place-based, organisational, or policy interventions related to transport, welfare, housing, crime, and employment—trialists may need to adapt the

**Table 4. Template for Intervention Description and Replication (TIDieR) Checklist<sup>40</sup>**

<b>Item Number</b>	<b>TIDieR Item</b>	<b>Definition</b>
1	Brief Name	Provide the name or a phrase that describes the intervention
2	Why	Describe any rationale, theory, or goal of the elements essential to the intervention
3	What: Materials	Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (e.g., online appendix, URL)
4	What: Procedures	Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities
5	Who Provided	For each category of intervention provider (such as psychologist, nursing assistant), describe their expertise, background, and any specific training given
6	How	Describe the modes of delivery (such as face to face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group
7	Where	Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features
8	When and How Much	Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity, or dose
9	Tailoring	If the intervention was planned to be personalised, titrated or adapted, then describe what, why, when, and how
10	Modifications	If the intervention was modified during the course of the study, describe the changes (what, why, when, and how)
11	How Well: Planned	If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them
12	How Well: Actual	If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned

TIDieR framework for their purposes; accumulated knowledge about reporting such interventions according to modified use of TIDieR may lead to the development of reporting standards for these more complex social interventions. As a starting point, the criteria for reporting the development and evaluation of complex interventions in healthcare (CReDECI) guidance may be useful to consult.<sup>41</sup>

Better communication of the content of interventions is of fundamental importance to understanding, replicating, and interpreting intervention impacts.<sup>27</sup> When describing the “what” of the intervention, or the actual procedures or techniques employed, authors should make use of established terminology, such as the Behavior Change Technique (BCT) taxonomy.<sup>42</sup> Authors should delineate any physical or informational materials that are part of the intervention or are part of the training and supervision of those delivering it. Information related to costs of these factors may also be helpful, if available. Authors should also indicate if there were any piloting or run-in periods of intervention delivery in the trial context prior to the actual start of the trial, noting any refinement or potential biases to trial methods or intervention implementation in the trial as a result.<sup>18</sup>

The format or mode of intervention delivery, the exposure to or amount of the intervention received, and the timespan of the intervention on offer are important to report.<sup>19</sup> The format or unit of delivery can vary quite substantially across social and psychological interventions, such as individual delivery online or in-person; group delivery, where the groups can be couples, families, schools, organisations, or whole communities; or even placed-based delivery to geographic units. Authors should note the unit of intervention and how participants may have been grouped or organised during delivery.<sup>12</sup> How to report exposure to and timespan of interventions will vary depending on the nature of the intervention. For instance, with those interventions involving sessions, the intended number, frequency, and length of each session should be reported, as well as the

overall duration of intervention implementation and the method of delivery (e.g., in-person versus online). In describing when the interventions occur, authors should note the timing of intervention activities relative to each other in addition to an intervention's overall duration. Authors may also consider tables or diagrams that delineate the sequence of activities in both the experimental and comparator interventions to clarify the contrast in programme offers.<sup>43</sup> For other types of interventions, such as place-based interventions, authors should report the duration of intervention implementation.

Authors should provide details about the specific providers in the trial. Information about providers can include the number of providers per trial group, their actual professional qualifications and training, expertise or competence with the interventions or area in general, training to deliver the intervention, and supervision of providers across groups. Issues to do with staff turnover should also be mentioned when significant. For large, multi-centre trials employing a large staff, authors should describe the minimal requirements for professionals to be enrolled in the trial as well as any additional training and supervision that they received. Authors should also discuss provider recruitment and selection procedures, as well as training and support of those delivering the intervention. Increasing evidence suggests that such procedures are part and parcel with the intervention.<sup>44</sup>

Authors need to be sure to define any comparator conditions to the same degree as the intervention group of interest. As the effect size for an intervention depends in part on the impact of techniques employed in the comparison group, the differences between the interventions in experimental and comparison groups needs to be clear in order to interpret the results.<sup>27</sup> Authors should avoid describing the “what” of comparison conditions merely as “treatment as usual” or “standard care”, as readers will not be clear about what exactly this entailed and thus about how to interpret the meaningfulness of results.<sup>18</sup> Terms such as

“treatment as usual” or “standard care” are often used when comparator conditions are left to the normal practices of addressing an issue in the setting where the trial takes place.

However, usual care is highly variable within and across a time and place: sometimes consisting of nothing and other times comprising an advanced, complex care system.

Authors should note what such usual care conditions might have actually involved in the trial, so that the reader can have an accurate understanding of the contrast between what is being offered in the experimental versus the comparator intervention.

*Extended CONSORT-SPI Item 5. Extent to which interventions were delivered and taken up as planned, including what they actually involved*

*Example:* “The intervention children adapted well to the Captain’s Log training program. Although only a few of the children had used a computer before, they understood task instructions in the local language of Luganda, used to orient them to the computer screen and trackball mouse. They also understood and responded correctly to the training items when guided through the various Captain’s Log training tasks. Intervention children usually became self-directive on the Captain’s Log tasks after one or two training sessions. By the third session, all intervention children were able to complete the training program on their own without instruction or close supervision... For the children with HIV undergoing CCRT (computerized cognitive rehabilitation therapy) intervention, eight children missed at least one scheduled session ( $M$  missed sessions = 2;  $N$  = 8). Because 304 of the 320 scheduled training sessions were completed as scheduled, this resulted in an overall adherence rate of 95%. Six children missed because schools were breaking for the holidays, one parent was missing for unknown reasons, and one child withdrew from the study after the first training session, leaving 32 children in the intervention group. Of the seven children remaining in the study, all eventually completed the required 10 CCRT sessions, although two of the children took a month to do so beyond the scheduled end date 5 weeks following their enrollment.”<sup>45</sup>

*Explanation:* Authors should provide evidence collected in the trial on how the intervention was actually delivered and taken up by participants. Intervention delivery refers to fidelity and related terms (e.g., integrity), whereas intervention uptake refers to adherence and related terms (e.g., compliance, reach). Information about the actual delivery of an intervention by providers and uptake by participants is essential to understanding what really occurred and therefore making appropriate inferences about the

results. As with the CONSORT 2010 Item 5 on interventions, details should ideally be provided for all comparator conditions as well. However, delivery and uptake of intervention conditions is rarely reported.<sup>17</sup>

Some details related to delivery and uptake may be reported in the methods section (e.g., techniques to assess fidelity), whereas other information may be provided in the results section (e.g., actual amount of intervention sessions attended). At a minimum, authors should aim to summarise the information about delivery and uptake that are key to interpreting the results of the trial. For example, researchers could indicate which components of an approach (e.g., affective expression, regulation, developing a trauma narrative, and processing for Trauma-Focused CBT) were most critical for affecting outcomes. For some higher-level interventions, “delivery” may involve facilitation of how the units of intervention worked towards intended goals, such as changes in policy or systems structures.

Details about the delivery of the interventions are particularly relevant when implementation differs from protocols or when it is not detailed within them.<sup>19</sup> Fidelity and adaptation may be in tension in some trials, as interventions can be provided in different ways depending on the immediate setting, organisation, and wider contextual factors. Clear descriptions of what needs to be standardised and what is free to be adapted should be provided. Modifications during implementation that were not intended or delineated in the intervention protocol should be distinguished from planned, systematic adaptations or tailoring. For example, TIDieR asks for information on pre-planned tailoring of the intervention as well as modifications made during the course of the study when reporting the actual delivery of the intervention compared to intervention protocols (see Table 4). For automated interventions (e.g., those online or using a computer programme), information about glitches in delivery and actual participant use may be relevant.

Participants are likely to respond differentially to social and psychological interventions. Consequently, social and psychological interventions typically show variability in the degree to which the intervention and control conditions were taken up as planned. Participant acceptability, satisfaction, and perceived value of the intervention may be reported to elucidate how participants respond to being offered an intervention. The amount of the intervention received by participants is also useful to report when known, such as number of sessions attended. Contamination or intervention diffusion across groups is one particular area of concern for social and psychological interventions, as the effect of an intervention may be reduced if content of the experimental intervention was exposed to the comparison condition.<sup>46</sup> However, exposure-related data may be particularly difficult to obtain or even irrelevant for place- or area-based interventions, as engagement with an intervention by particular individuals is not assessed. Where available, trialists evaluating place-based interventions should look to present information on geographic reach of the intervention (i.e., the extent to which a target audience comes into contact with the intervention).<sup>13</sup>

Authors should describe any procedures in place to enhance fidelity of implementation to the intervention design, as well as any methods—quantitative or qualitative—for investigating intervention delivery and uptake. Reporting on process evaluations that investigate intervention implementation can help to elucidate this information. For example, therapist or provider competence investigated in process evaluations of psychotherapy—such as the quality of provision in addition to fidelity to a manual—may be helpful to report. Full details of a process evaluation may be too much to include a main trial report. Given word limits in journals, further information may be available in online supplements, sibling publications, or other outputs.

*Extended CONSORT-SPI Item 5. Where other informational materials about delivering the intervention can be accessed*

*Example 1:* “To assist in dissemination and further evaluation of the developed workshop, the Powerpoint presentation, screening measures, and all protocols are freely available from the corresponding author.”<sup>47</sup>

*Example 2:* “The protocol used in the ACT (acceptance and commitment therapy) for OCD (obsessive-compulsive disorder) investigation is the same as that used by Twohig et al. (2006a)... The protocols are available at <http://www.contextualpsychology.org> or from the first author.”<sup>48</sup>

*Explanation:* Readers benefit from understanding what exactly was done in each intervention and the materials used to facilitate delivery. However, a full description of the intervention that allows it to be replicated is generally not possible in a journal article, given space restrictions. As such, authors should indicate where informational materials about implementing the intervention can be located.

Some interventions may have a full manual that can be referenced. For example, *Addiction* has instated a policy whereby the full manual for experimental and comparator interventions in every RCT manuscript must be available as supplementary material or in another accessible location.<sup>49</sup> If proprietary, a reference to where such materials can be purchased should be provided. More complex interventions that allow greater flexibility in implementation may have intervention protocols or plans that should be supplied. For policies, a statute or legislation could be referenced. Other sources may include video recordings that demonstrate how to deliver an intervention, websites with further materials, or online depositories of intervention delivery materials.<sup>27</sup> Authors may also wish to provide materials on how to train providers, when relevant.<sup>18</sup> Professional websites tend to be impermanent, so use of a sustainable online service should be used if URLs are provided. Those employing “brand name” interventions should note if any adjustments were made to referenced materials, and should be sure to cite the proper materials if more than one version of a brand name intervention exists.

*Extended CONSORT-SPI Item 5. When applicable, how intervention providers were assigned to each group*

*Example:* “Facilitators were assigned to conduct groups on the basis of (a) academic or clinical training experiences that enhanced a facilitator’s ability to lead either CB (cognitive–behavioural) or supportive expressive groups; (b) preference for the therapeutic approach (CB or supportive expressive); (c) time availability to conduct a group on a weekly basis for 6 weeks; and (d) gender (the gender of at least one of the facilitators matched the gender of group participants).”<sup>50</sup>

*Explanation:* As part of the intervention, differences in the providers between the experimental and comparator are needed to understand the intervention contrast and thus interpret the effect size. For example, some authors report randomising providers to different conditions to help ensure that important factors like self-selection, expertise, or allegiance are not involved in assigning providers to trial groups. Authors should report whether the same people delivered the intervention(s) and comparator(s), or whether providers were nested within conditions. Where applicable, the caseload for each provider should also be given (e.g., teachers had classes of 15-20 students on average).

### **3.6 Methods: Outcomes**

*Item 6a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed*

*Example:* “The assessments described below were administered at four time-points: baseline (pretreatment), post-treatment (2 months) and at 3 and 6 months post-treatment. All assessments were sent out and processed by a research assistant who was blind to treatment condition. If assessments were late in returning, a reminder letter was sent out, followed by a telephone call. Two primary outcomes were included. The first, the Subject’s Global Assessment (SGA) of Relief is used frequently in treatment trials to identify IBS (irritable bowel syndrome) responders to therapy. Participants rate their relief from IBS symptoms on a scale of 1 to 5 ranging from ‘completely relieved’ to ‘worse’. Scores are dichotomized so that patients scoring from 1 to 3 are considered responders and those from 4 to 5 non-responders. Because IBS is a cyclical illness, responders in this trial were defined as those who scored between 1 and 3 at all three follow-up time-points. Patients who returned no follow-up questionnaires were rated as not relieved. Those who returned one or more questionnaires were considered responders if they had no ratings of 4 or 5 at any time-point. The other primary outcome was the Irritable Bowel Syndrome Severity Scoring System, which measure the severity of pain, distension, bowel dysfunction and quality of life/global well-being. A decrease of 50 points on this scale has

been identified as a clinically significant change in symptom severity. Secondary outcomes included the Work and Social Adjustment Scale (WSAS) and the Hospital Anxiety and Depression Scale (HADS). The WSAS is a measure of quality of life that addresses the degree to which a patient's illness impacts on work, personal relationships, home management, and social and leisure activities. The HADS is a commonly used self-report instrument for detecting depression and anxiety in patients with medical illnesses. The IBS-SS, WSAS and HADS were all used in a previous primary-care IBS CBT (cognitive-behavioral therapy) trial, which allowed the data from this trial to be compared with earlier work.<sup>20</sup>

*Explanation:* Outcomes refer to those measures collected after random assignment.

Authors should completely define all primary and secondary outcomes declared in the trial report so that the reader can judge the relevance and suitability of variables and the methods used to measure them. Namely, an interested reader would want to be clear about the outcome domain being investigated (e.g., school attendance), that these domains appropriately fit the hypothesis of how the intervention is supposed to work, that the methods used to measure this outcome domain are valid and reliable to detect change in the trial context, and which measures were chosen in advance. Authors should also define measures used as covariates in analyses of outcome variables, and ideally have some indication of measures collected but not reported in a manuscript—for example, by referencing a trial registration number (Item 23) or trial protocol (Item 24).

Social and psychological interventions may have multiple outcomes that are seen as “primary”, given that they are often designed to address several outcomes. Still, the outcomes delineated as primary in the report should not differ from those delineated as primary during the design of the trial. These primary outcomes should follow from any theory of change discussed above, with a match between conceptual thinking about how the intervention works and measurement models that best capture what variables to measure, how to measure them, and when they should be measured. Specifying primary outcomes eases interpretations of results where the multiplicity of outcome measures increases the chances of concluding the intervention is effective.<sup>18</sup>

Complete definition involves the nature of the measure, the format used to collect the data, and the timing of all measurements. Rather than using general terms such as “well-being”, authors should be specific about the measurements taken, particularly for complex variables like organisational capacity or social capital. The methods used to collect the data—such as questionnaires, interviews, behavioural assessments, or observations—should be clear for each outcome. Psychometric properties of any questionnaires related to their reliability and validity in the trial context should be addressed, either via references to previous validation studies or discussion of these properties in the trial, particularly if the instrument was created or adapted for the study.<sup>19</sup> If validation studies are cited, authors should note key similarities and differences between the context of the validation study and that of the trial. If possible, it is preferred that authors provide information on how to access questionnaires—ideally by depositing copies of modified or new questionnaires in a repository or online supplement. Such information would be useful to allow for replication, though the proprietary nature of some questionnaires and the resources of some journals may serve as barriers. Authors should note whether choice of length of follow-up was due to pragmatic reasons (e.g., funding resources) rather than devising theoretical rationale post-hoc. It should also be clear from reports whether follow-up periods are relative to baseline or the end of the intervention period (see Item 14a).

Should an outcome variable be measured differently between trial groups (e.g., at different points in time, using a different format), this should be reported. Techniques to enhance the quality of measurements may also be reported, such as the training of outcome assessors or using multiple observations (e.g., parent and teacher reports of a child behaviour). Authors may also wish to note the potential confounding of measurement methods and the impact of the interventions on outcomes. For example, extensive baseline data collection amongst comparator participants receiving “no intervention” may actually

serve as an intervention, or the outcome measures chosen may “teach to the test” of the experimental intervention. In social and psychological intervention trials, outcomes may also be mediators of the intervention or measures of intervention implementation, such as fidelity to intervention protocols by providers. When reported, these should also be completely defined. If done, authors should report methods used to enhance the quality of measurements, such as training procedures for data collectors or use of multiple observations of the same outcome.<sup>19</sup>

*Item 6b. Any changes to trial outcomes after the trial commenced, with reasons*

*Example:* “Because of the relatively low rate of apparent underreporting by participants relative to collaterals, and the small percentage of participants for whom collateral data were available, participant self-reports were used in the analyses of alcohol use.”<sup>39</sup>

*Explanation:* Outcome measures ideally should be described in a trial protocol that was written and published prior to data collection and analysis. However, methods may be adjusted after the trial has begun due to unforeseen circumstances. Authors should note any changes to the outcomes in a trial, including removing or adding outcome measures altogether, or changing a way that a particular outcome is being measured, such as shifting from an interview to a self-report format. Changing pre-defined outcomes once an RCT has begun requires proper explanation.

### **3.7 Methods: Sample Size**

*Item 7a. How sample size was determined*

*Example:* “The required sample size was calculated for patients receiving the intensive intervention (<95% timing adherence), using data obtained during the pilot-study. Based on our primary outcome timing adherence, two-tailed testing with  $\alpha = .05$  and a power  $1-\beta = 0.90$ , a standardized effect size (Cohen’s  $d = 0.70$ ), and an anticipated dropout of 15%, per Group 50 participants with <95% timing adherence at baseline were required. However, because this power calculation does not account for the power increase resulting from including baseline adherence as covariate, a sample size of 40–50 patients with <95%

adherence per condition was considered satisfactory. Also using the pilot-study data, we estimated that 50% of the patients would score <95% adherence at baseline. Based on these calculations, the objective was to include 160–200 patients in the study.”<sup>51</sup>

*Explanation:* Authors should indicate the intended sample size for the trial and how this sample size was determined. Authors may indicate that the sample size was determined either by methodological reasons—such as an *a priori* sample size calculation—or for pragmatic reasons, like the number of participants available in a setting or the number of participants that trialists had the resources to recruit. When an *a priori* sample size calculation or power analysis was conducted, authors need to report the anticipated effect estimate and its source. Estimates for the same intervention effect may vary depending on the source; an effect size based on a single previous trial with questionable applicability to the current trial’s context may differ from an effect size from a previous systematic review or one based on a practically meaningful difference. Though sample size calculations are typically performed for the primary outcome, the ability to detect desired effects on secondary outcomes from the estimated sample size should also be noted. It is important to note whether the desired sample size was recruited and then retained at each follow-up. If an *a priori* sample size calculation was not performed, authors should not present a post hoc calculation, but rather report the genuine reason for the sample size as well as the actual power to detect an effect for each result from the sample size obtained (Item 17). Several resources exist for sample size planning before a trial, such as those offered by the Australian National Statistical Service.<sup>52</sup>

*Item 7b. When applicable, explanation of any interim analyses and stopping guidelines*

*Example:* “An independent data monitoring committee reviewed unblinded data for safety after the first 1000 women in the study had given birth. In response to a lower than anticipated attrition rate, we stopped recruitment when 1748 had been randomly assigned.”<sup>53</sup>

*Explanation:* Interim analyses involve examination of available outcome data prior to the completion of data collection. Of particular concern is the use of predetermined stopping guidelines based on detections of statistical significance, which is at high risk of stopping trials due to chance findings. Stopping guidelines serve as a tool for deciding whether the trial should be ended early, such as ethical concerns related to harm being clearly and significantly experienced by one trial group. Data monitoring committees typically review outcome data when these procedures are in place. Though not common in social and psychological intervention RCTs, interim analyses and stopping guidelines are important to report when existent, as some trialists do use them to decide when to cease recruitment.<sup>19</sup> If not part of the trial, authors can simply state “no interim analyses of the data were planned or conducted.”<sup>54</sup>

### **3.8 Methods: Randomisation—Sequence Generation**

*Item 8a. Method used to generate the random allocation sequence*

*Example 1:* “Random allocation was managed by the study statistician (S.M.C.) using computer-generated random numbers.”<sup>55</sup>

*Example 2:* “Patients were randomly assigned to a brief and an intensive intervention condition using a random number generator.”<sup>56</sup>

*Explanation:* Randomisation underpins the strength of RCTs in yielding reliable estimates of intervention effects. Use of a method that generates a truly random sequence of numbers is needed for this promise to be met. The author should report the actual method to generate the random allocation sequence, such as a computer, random numbers table, or coin toss. Reporting this information helps the reader assess bias, as not all methods used are truly random.

*Item 8b. Type of randomisation; details of any restriction (such as blocking and block size)*

*Example 1:* “In each scheme, a blocked randomization scheme, with blocks of 30, was used to yield a balanced allocation of participants to the three treatment groups.”<sup>39</sup>

*Example 2:* “Randomization to HOPES or TAU within each site was stratified by diagnosis (mood disorder or schizophrenia spectrum) and gender.”<sup>57</sup>

*Explanation:* Blocking entails creating groups or blocks of participants, and creating different random sequences of numbers within each of these blocks. It is typically used to ensure equal numbers of participants in each condition when the sample is expected to be small.<sup>19</sup> Stratification involves creating different random sequences of numbers across important strata of participants, such as one sequence for males and another for females, to ensure that trial groups are balanced on these characteristics.<sup>19</sup> Stratification is typically done for factors that are likely modifiers of the intervention effect. Authors should provide details of any such methods, including the size of blocks and the variables used for stratification.

### **3.9 Methods: Randomisation—Allocation Concealment Mechanism**

*Item 9. Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned*

*Example:* “The randomization list was transferred to a sequence of brown envelopes by writing the sequence of treatment names on the inside of the envelopes, which were then sealed. The sequence of envelopes was then ‘cut’ by taking approximately the first half of the envelopes and placing them at the end of the sequence so that no person involved in the trial would know the starting point of the randomization sequence and to preserve allocation concealment. The envelopes were then numbered.”<sup>54</sup>

*Explanation:* Allocation concealment relates to the issue of whether trial staff enrolling participants in a study are aware of the sequence of assignment. Authors should report methods to keep allocation concealed, such as the use of opaque and sealed envelopes, particularly when the same person is involved in generating the random allocation

sequence, recruiting participants, and assigning participants to interventions.

### ***3.10 Methods: Randomisation—Implementation***

*Item 10. Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions*

*Example:* “Patients were screened by our research staff, and those who met inclusion criteria were told that they would become eligible for the study after they had completed at least 2 weeks of attendance in the program... The randomizations were generated by one of the authors (KGL) and stored in envelopes, and the sequence was concealed until the treatment assignment was provided to the participant. Participants were notified of their treatment assignment by one of the research staff at the end of the baseline assessments.”<sup>58</sup>

*Explanation:* The role of trial staff in generating the random sequence, enrolling participants, and assigning participants to interventions should be described. Where possible, authors should note the staff who carried out each of the above procedures. Division of this labour is preferred to further avoid potential selection biases when those enrolling participants may have knowledge of the randomisation sequence as well as prognostic characteristics of participants.<sup>18</sup> The concern with these processes overlapping may not be as pertinent with group- or place-based interventions, where all clusters are recruited first and then randomised at the same time. The staff used to enrol participants can also have implications for recruitment rates and subsequent applicability of trial results, particularly if there are any important differences from standard referral or recruitment procedures in the setting. Those enrolling participants may also have a conflict of interest, for instance if the results of the trial could impact their employment moving forward.

### ***3.11 Methods: Awareness of Assignment***

*Item 11a. Who was aware after assignment to interventions (for example, participants, providers, those assessing outcomes), and how any masking was done*

*Example 1:* “The therapists, participants, and graduate research assistants were not blind to intervention group assignment. However, it would not be possible to blind the therapists or the participants because of the nature of the intervention. With respect to the research assistants, all assessments were conducted online, reducing the likelihood of biases due to nonblinding.”<sup>59</sup>

*Example 2:* “Although participants and clinicians delivering the treatment could not be blinded to treatment assignment, assessors and clinicians conducting outcome assessments were blinded. In addition, participants were instructed at their follow-up assessment interviews not to reveal their treatment assignment.”<sup>60</sup>

*Explanation:* Masking refers to methods of withholding information about assigned interventions *post-randomisation* from those involved in the trial, when knowledge of this information could influence their behaviour in a way that would later prove integral to interpreting the results. The term “masking” is increasingly preferred instead of “blinding”, which uses a term for a disability in a metaphoric and potentially offensive sense;<sup>19</sup> this guideline encourages authors to use “masking”. The importance of masking relates to biases that may hinder the internal validity of a trial after random assignment. In medical trials, authors of RCTs typically indicate whether participants, those delivering the interventions, and those assessing outcomes were kept unaware of assignments of participants to trial groups.<sup>14</sup> However, masking of providers and participants is generally not possible in social and psychological interventions, as awareness of what is being delivered and received is an explicit and even necessary aspect of the intervention rather than something to be separated from it.<sup>61</sup> Given such infeasibility, then, lack of masking participants and providers is not a methodological weakness per se of specific social and psychological intervention trials, but rather has implications for the interpretation of trial results in this area generally.

If possible, authors should note the method for masking providers or participants. Otherwise, authors should note that blinding of participants and providers was not possible and then describe methods, if any, that were employed to minimise the potential for the

performance and expectancy biases that masking aims to address. These can include keeping participants unaware of trial hypotheses, measuring participant and provider expectations of benefit at baseline, and measuring providers' allegiance to the intervention that they are delivering. Of particular importance is whether the same providers were used across conditions, and whether providers were potentially vested in the experimental or comparator intervention. The existence of any important biases, such as demoralisation effects of being randomised to a comparator group, should be noted if suspected or known.

Compared to providers and participants, masking of outcome assessors is possible in social and psychological intervention RCTs, and is important given the often "subjective" nature of many measurement methods in this area. Information about methods to mask assessors is particularly important for outcomes assessed via interview or observation. However, authors should note any limitations or issues related to masking outcome assessment, considering that different informants or sources of outcomes may be more or less susceptible to bias. For example, self-reported outcome measures generally cannot be masked, as participants are aware of the interventions they receive; these types of outcomes are common in this area,<sup>17</sup> particularly given the recent emphasis on participant-centred outcome research. Moreover, some trials may not have "outcome assessors" in the traditional sense, for example by using administrative data or more objective measures (e.g., blood tests), where masking is less of an issue. In other circumstances, masking may have been attempted at the outset, yet over the course of follow-up, factors may have led masking to be broken. If trial resources allowed, authors should report whether those analysing the data were able to be masked from trial group assignment, particularly if the principal investigator or those with a competing interest are analysing the data—though the resources to do so may not be common.

*Item 11b. If relevant, description of the similarity of interventions*

*Example:* “The attention control condition was led by the same interventionists who led the hypnosis intervention sessions. However, the interventionists did not lead the attention control patients in imagery, relaxation, or even simple discussion. Rather the interventionists allowed patients to direct the flow of the conversation and provided supportive and empathic comments according to standardized procedures.”<sup>62</sup>

*Explanation:* Differences between the experimental and comparator interventions is essential to understand the contrast in programme offers and thus interpret the ultimate effect sizes generated. Compared to the section on interventions above, this item particularly focuses on dissimilarities between the interventions that could lead to differences in the performance and expectancies biases that masking providers and participants would have aimed to minimise. Social and psychological interventions, while sharing many features, likely differ in important ways that lead participants in a trial to react to awareness of what they are receiving and what they could have received in another trial group, particularly where participant preferences for one form of intervention are present and known.<sup>18</sup> In addition to intervention content, it is important to know if *context* differs between intervention and control groups in important ways, because of the potential for features of the context to be necessary auxiliary factors for intervention techniques. Differences in potential co-interventions by trial group are also important to report if relevant and known.

### **3.12 Methods: Analytical Methods**

*Item 12a. Statistical methods used to compare groups for primary and secondary outcomes*

*Example:* “Repeated measures analyses were conducted with generalized estimating equations (GEEs) to test group differences in the odds of being abstinent at 1, 4, 16, and 26 weeks post assigned quit date for both the full randomized sample (n = 68) and the subsample attending at least one treatment session (n = 42). Included covariates were gender, nicotine dependence, BDI-II symptoms, and current income—all of which are commonly linked to poor cessation outcomes. We also included a linear effect of time.”<sup>63</sup>

*Explanation:* Whereas Item 6 involves what the data in the trial are, Item 12 relates to the strategies or methods used to analyse data to compare groups on primary and secondary outcomes of interest. Clear statistical reporting allows the reader to replicate analyses with access to the data (Item 17) as well as better understand the reliability of the data.<sup>18</sup> The authors need to state whether the analyses were *a priori* or decided after data was collected. Any changes to pre-determined data analysis plans should be noted. For complex correlated data, such as when units are nested within clusters, authors need to account for the dependence of data within clusters in their analyses.<sup>8,12,19</sup> Authors should also note the smallest unit (e.g., individuals, groups) that is being analysed to assess intervention effects, and, if this unit of analysis differs from the unit of assignment, the analytical methods they used to account for this issue, such as multi-level modelling.<sup>12</sup> For complex or novel statistical techniques, it may be helpful to provide the software and code used. Procedures for any changes to the raw data, such as transformations or deleting cases, should also be stated along with problems about statistical assumptions or data distributions that they aimed to address (e.g., estimation problems, anomalous data points), so that readers can assess their impact on the validity of the findings.<sup>19</sup>

*Extended CONSORT-SPI Item 12a. How missing data were handled (e.g., complete case analysis, simple imputation, multiple imputation), with details of any imputation method*

*Example:* “Core analyses used an intent-to-treat, listwise missing value procedure, wherein we analyzed all participants who had data for all waves (n = 251 [comparison n = 120, intervention n = 131]; 80% of participants). We determined these analyses to be appropriate by testing whether data were missing completely at random, meaning that the probability of observing a case is independent of the values of any independent or dependent variable. We compared these results with analyses in which we imputed missing values among participants with at least one follow-up wave (n = 297; 95% of participants). For continuous outcomes, we used the multiple imputation procedure from SAS; because the data were non-monotonically missing, we used the Markov Chain Monte Carlo procedure. We used all available data regarding demographic status, psychosocial variables, UAI (unprotected anal intercourse), and transmission risk partners to impute missing values on risk outcomes. Missing data correction for binary measures used the previous wave value. This was very conservative because most missing data were at the 6-month follow-up and were replaced by the baseline value.”<sup>25</sup>

*Explanation:* Missing data is a common problem in social and psychological intervention RCTs for a host of reasons, such as participant drop-out, inability to locate participants at a particular follow-up point, or issues with measures used.<sup>19</sup> For any missing data problems, authors need to report the frequency or percentages of missing data, empirical evidence or theoretical arguments for the authors' judgments as to why the data are missing, and any methods for addressing missing data that were used.<sup>19</sup> Many methods are used to provide surrogates for missing data, particularly in conducting intent-to-treat analyses in RCTs (Item 16). It is important to note how trialists handled missing data in order to determine the likelihood of biased estimates of intervention effects. Certain methods for dealing with missing data are more appropriate than others, as some may increase risk of bias (e.g., last observation carried forward for long-term follow-ups). For imputation methods that aim to estimate the value of missing data based on values of other data in the dataset, the variables used for imputation and the number of imputations performed would be useful to report, as well as any advanced software procedures for executing the imputations, to allow for replication. Authors should discuss any maximum likelihood procedures if used. Authors should also report the amount of missing data and discuss any assumptions about the "missingness of data": i.e., how they judged data to be missing completely at random, at random, or not at random.<sup>64</sup>

*Item 12b. Methods for additional analyses, such as subgroup analyses and adjusted analyses*

*Example:* "The moderator variable, High/Limited Prevention Opportunity was also dummy coded (1 = High Prevention Opportunity; 0 = Limited Prevention Opportunity). The High Prevention Opportunity (HPO) subgroup was composed of women who met all 3 of the following criteria: first-time mother, under the age of 19 years, and randomly assigned at a gestational age of 30 weeks or less. The Limited Prevention Opportunity (LPO) subgroup included all other mothers, that is, those who met none or up to 2 of these criteria, but not all 3 of these conditions.... Logistic regression analyses were used to examine the effects of HFNY on each of the 9 dependent measures at Year 3 (i.e., positive parenting, harshness, and role reversal for each of the 3 tasks).... An interaction term that represented

the combined effects of the treatment and HPO/LPO subgroups was formed by multiplying the HPO/LPO subgroup variable by treatment group assignment. SAS 9.0 was used to construct the multivariate models.”<sup>65</sup>

*Explanation:* Authors should discuss any additional analyses beyond those looking at primary and secondary outcomes, indicating which were pre-specified and which were exploratory after the data were collected. Highlighting analyses as additional or exploratory, when applicable, makes these results less likely to interfere with interpreting the trial’s main findings, particularly when a multitude of tests could lead to inflated statistical error rates.<sup>19</sup> Adjusted analyses typically involve the use of covariates to increase the precision of statistical tests or to adjust for potentially meaningful baseline differences, even though these are due to chance in properly randomised trials.<sup>19</sup> For adjusted analyses, authors should report the statistical procedures and covariates used. Sensitivity analyses may also be reported, e.g., those assessing the impact of using different methods for handling missing data.

Additional analyses may include testing the mechanisms or causal pathways through which the intervention was hypothesised to work, for example via a mediation analysis of therapeutic alliance in a psychotherapy trial.<sup>66</sup> Authors should also report the methods for conducting sub-group analyses that investigate potential inequities caused by the intervention. Though likely to be fully reported in another paper, qualitative methods embedded in trials to investigate implementation processes, contextual influences, and unanticipated outcomes may also be described if done.

### ***3.13 Results: Participant Flow***

*Item 13a. For each group, the numbers randomly assigned, received intended treatment, and were analysed for the primary outcome*

*Extended CONSORT-SPI Item 13a. Where possible, the number approached, screened, and eligible prior to random assignment, with reasons for dropout*

*Example:* See Figures 2, 3, and 4.

*Explanation:* Authors should report any information they have about the total number of participants at each stage of the trial, with reasons for participants missing at each stage.

Attrition at any stage—from approaching participants to outcome assessment—can significantly impact the internal validity of a study, its generalisability, or both.<sup>19</sup>

Information about the flow of participants through each stage of a trial—approach, screening, enrolment, assignment, intervention exposure, follow-up, and analysis—is thus essential for understanding both the sampling frame of the study and the rigour of the data ultimately collected. Approach involves the number contacted for interest in participating in the study. Screening and enrolment involve the number screened for eligibility, found to be eligible or not, those eligible declining to be enrolled, and those ultimately enrolled. Assignment and intervention exposure involve the number assigned to each trial group and who actually received each intervention. For each outcome assessment, follow-up and analysis involves the number in each trial group who completed the assessment, did not complete the assessment and reasons why not, and were ultimately included in analyses.<sup>12</sup>

Though some opt to report such information in the text, a flow chart is strongly recommended to provide a succinct summary of this information, indicate overall attrition from the trial and differential attrition across trial groups, and ensure that none of the requested information is forgotten. It also helps to clarify the relative timing of each of the key study procedures (see Item 14a). When the information is available, authors should distinguish the number of participants who discontinued receipt of the intervention but still provided outcome data from those participants who dropped out of the trial altogether. Some trials may not have accurate information about the number of participants approached and screened because the researchers didn't collect the data, or it wasn't possible to collect the data (e.g., universal prevention interventions, trials using public

announcements as part of recruitment). In these instances, approximate information about the sampling frame may be reported instead, such as reporting the number of online Listservs sent announcements about the trial.

*Item 13b. For each group, losses and exclusions after randomisation, together with reasons*

*Example:* “Two of the 502 BMI (brief motivational intervention) participants were administratively dropped from the study when it was discovered after randomization that they began working at the university survey research center collecting data for the study.”<sup>70</sup>

*Explanation:* With similar rationale to the Item 13a above, authors should report participant attrition or data exclusion by the research team per trial group at each follow-up point, with reasons for attrition or exclusion as relevant.<sup>18</sup>

### **3.14 Results: Recruitment**

*Item 14a. Dates defining the periods of recruitment and follow-up*

*Example:* “We recruited participants from April 2006 to January 2008 using radio, web-based, and newspaper advertisements.... Smoking was assessed at 1 week, 4 weeks (end of behavioral treatment), 16 weeks, and 26 weeks post assigned quit date.”

*Explanation:* Authors should define the time frame of the trial from recruitment procedures through to the last outcome assessment. Where possible, authors should aim to provide a full schedule of trial activities, including recruitment practices, baseline assessment, period of intervention delivery, and schedule of assessments. Specific dates can be particularly important to know when certain aspects of the trial or data are tied to particular circumstances, secular trends, or historical events.<sup>18</sup> For example, evaluators of a parenting programme in Louisiana, USA in the fall 2005 had to significantly modify the model of the programme, as Hurricanes Katrina and Rita hindered the ability of the child welfare system to deliver services.<sup>71</sup>

Authors should also be clear whether various waves or phases of recruitment occurred during the recruitment period. For multi-faceted interventions that involve many different strategies, it is particularly helpful to clearly articulate when particular aspects of the intervention have actually started being delivered in relation to other features of the trial, such as follow-up. Authors should note whether baseline assessment occurred before or after randomisation: it is important to define follow-ups relative to a certain study procedure (e.g., baseline, end of intervention), as “4-week follow-up” could mean very different things if meant after baseline or after a 2-month intervention.

*Item 14b. Why the trial ended or was stopped*

*Example:* “Mailings were stopped once the study sample size was achieved because of study time lines and budget constraints.”<sup>72</sup>

*Explanation:* Authors should note the status of the trial at the time of publication. If a trial ended or was stopped, authors should note whether the reason was part of an *a priori* decision plan (e.g., sample size reached and pre-determined follow-up period completed), or an unplanned decision in response to what was happening in the trial. It is particularly key to mention if there were ethical issues or insurmountable challenges related to implementation (e.g., could not recruit enough participants), or if the trial was stopped early perhaps due to a positive effect, which may lead to exaggerated effect sizes. For trials with long-term follow-up, authors may report interim findings and indicate that outcome data are still being collected. In addition, some field trials may experience interruption for a time (e.g., a natural disaster or security crisis) and then continue at a later date.

**3.15 Results: Baseline Data**

*Item 15. A table showing baseline demographic and clinical characteristics for each group*

*Extended CONSORT-SPI Item 15. Including socioeconomic variables where applicable*

*Example:* See Table 5.

*Explanation:* Authors should provide a table for data collected at baseline by each trial group. Ideally, all data collected at baseline should be described rather than a selective presentation of baseline data. Essential are potential prognostic variables and pre-test measures of outcome variables. Topic-specific socioeconomic information related to equity concerns are particularly important to report for social and psychological intervention trials. The PROGRESS-Plus framework is a useful conceptual tool for understanding different mechanisms by which inequities may be propagated by an intervention, and thus which socioeconomic indicators to report.<sup>74</sup>

### **3.16 Results: Numbers Analysed**

*Item 16. For each group, number included in each analysis and whether the analysis was by original assigned groups*

*Example:* “The primary analyses used all available follow-up data and compared participants in their randomized groups, irrespective of the intervention they received. The sensitivity of the primary analyses was assessed including baseline school attendance, using a per protocol analysis (excluding three participants in the psycho-education group, two of whom did not fulfil criteria for CFS (chronic fatigue syndrome) and one who received 13 sessions of CBT (cognitive behavioural therapy)) and multiple imputation as an alternative method for handling missing data.”

See also Table 6 for number of participants for each analysis.<sup>54</sup>

*Explanation:* Identifying the number included in each analysis helps the reader know the completeness of data for an analysis and perform secondary analysis of the data. This information is particularly helpful when the number may be different across analyses, for example due to different follow-up rates across time-points and measures.<sup>12</sup> High attrition rates can have serious implications for the validity and generalisability of results.

Authors should be clear whether each reported analysis is intent-to-treat or a per-protocol. Intent-to-treat analysis refers to including all participants in the analysis no matter their level of adherence to the intervention protocol.<sup>19</sup> As such, the intervention

effects do not necessarily reflect the impact of receiving the intervention, but rather of being assigned to and offered a particular intervention. A per-protocol analysis refers to including in the analysis only those participants who received and experienced the interventions as intended. While this analysis provides an impact of the intervention effect, it breaks randomisation and therefore risks introducing confounds by comparing two groups that systematically differ on a variable other than the intervention assigned. As such, intent-to-treat analyses are recommended as the primary analysis for trials (though authors may need to deal with missing data to conduct these analyses), with the option to include per-protocol analyses as well (Item 12a).

### ***3.17 Results: Outcomes and Estimation***

*Item 17a. For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)*

*Examples:* See Tables 7 and 8

*Explanation:* Authors need to report the summary descriptive data for each outcome by trial group in order to facilitate subsequent re-analysis of the data. For continuous outcomes, this would include the mean, standard deviation, and number analysed for each outcome; for dichotomous outcomes, this would include the proportion of participants with and without the event for each group.

An effect size indicates the magnitude and direction of the effect of the intervention in comparison to the control group. Common effect size measures include the mean difference for continuous outcomes and odds or risk ratios for dichotomous outcomes. Authors may provide standardised mean differences, such as Cohen's  $d$  and Hedge's  $g$ , that provide scale-free measures of effects, though unstandardised measures may provide more interpretable and meaningful information by using a scale with clear cut-offs for clinical significance, such as grade point average. Authors should also consider statistical

dependence amongst effect sizes or adjustments for multiple comparisons. Estimates of the precision of the effect size should be provided, even if there is no statistically significant effect.<sup>19</sup>

*Extended CONSORT-SPI Item 17a. Indicate availability of trial data*

*Example 1:* “Data sharing: Patient level data and the statistical code are available from the corresponding author at s.priebe@qmul.ac.uk. Consent for data sharing was not obtained but the presented data are anonymised and the risk of identification is low.”<sup>75</sup>

*Example 2:* “Data sharing: No additional data available.”<sup>69</sup>

*Explanation:* Indicating the availability of trial data matches scientific norms of complete and transparent sharing of information. It is becoming more common for researchers to have an interpretable version of their data set and analysis code, and provide a copy of their raw data in an online supplement or online archive. Calls for data sharing and open access to raw data represent a powerful movement in the scientific community at the time of writing. The AllTrials campaign is one example that could benefit the social and psychological intervention community. Data sharing agreements by funders such as the Wellcome Trust, US National Institutes of Health, and UK National Institutes of Health Research are others. However, data protection laws or embargo policies may differ regionally; any restrictions imposed on making the trial data available should be noted, for example those made by the funder or ethics committees.

If trial data are available to readers, the authors should indicate how to obtain it. Where possible, it is preferred that data be stored in a sustainable, publicly accessible database over emailing a corresponding author, as authors can change institutions and over time may be difficult to locate. Authors should also provide any codebooks needed to actually use the data. When relevant, authors should also disclose any previous publications, conference papers, or even dissertations where data from a trial report can be

found, such as analyses of baseline data or the first follow-up time point.<sup>19</sup> In addition to indicating where the reader may obtain more data on the trial, this information helps prevent misunderstandings that multiple papers represent multiple datasets.

Some analyses in RCTs are very complex, requiring more than just summative data to meta-analyse or appraise. Rather than providing, for example, extensive variance-covariance matrices, making data available to those outside of the trial team can assist use of the data in secondary analyses as well double-checking the analyses performed in the trial. If the raw data is not available, the authors should indicate the availability of associated variance-covariance (or correlation) matrices for multivariable analytic systems, such as regression analyses, structural equation modelling analyses, and hierarchical linear modelling.<sup>19</sup> These are important for secondary analyses of data.

*Item 17b. For binary outcomes, presentation of both absolute and relative effect sizes is recommended*

*Example 1:* “Based on observed data, 24.3% (N=27/111) of the patients in the CBT (cognitive behavioural therapy) condition and 21.3% (N=26/122) in the psycho- dynamic therapy condition met the remission criterion at the posttreatment assessment. . . . At the posttreatment assessment, the odds ratio was 0.82 (95% CI=0.45–1.50), indicating that remission rates did not differ significantly.”<sup>76</sup>

*Example 2:* See Table 9.

*Explanation:* Odds and risk ratios express the relative strength of effect of the intervention compared to the control group, whereas absolute effect sizes like the risk difference indicate the actual difference in the observed risk of events between experimental and control interventions.<sup>77</sup> The actual event rates and the number of participants providing data for them in each trial group are important to provide in order to calculate both absolute and relative effect sizes.

### **3.18 Results: Ancillary Analyses**

*Item 18. Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory*

*Example:* “No main effects of the program were found for harsh parenting and role-reversed parenting at Year 3 for any of the tasks. However, as shown in Table 4, membership in the HPO (High Prevention Opportunity) subgroup moderated the relationship between treatment group assignment and observations of harsh parenting at Year 3... Table 4 presents the adjusted means produced by models including the interaction terms. As shown, HFNY (Healthy Families New York) mothers in the HPO subgroup were less likely to engage in any harsh parenting during the Puzzle Task (5.3% versus 21.5%) and the Delay Task (5.3% versus 23.8%) than their counterparts in the control group. Subsequent analyses revealed significant main effects of the intervention within the HPO subgroup for both the Puzzle Task ... and the Delay Task ...; whereas, main effects of the intervention were not detected within the LPO (Limited Prevention Opportunity) subgroup.... In addition, no moderator effects were found for role reversal.”<sup>65</sup>

*Explanation:* As with primary and secondary outcome data, the results for any ancillary analyses should be reported as well. If doing adjusted analyses, authors should always provide unadjusted summary data, or make the raw data available (see Item 17 above).

### **3.19 Results: Harms**

*Item 19. All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)*

*Example:* “Regarding major incidents, two suicide attempts, two accidental deaths related to psychotic symptoms, a serious fight where the patients sustained serious injuries and three patients initiating substance abuse were recorded in the control group. In the family intervention group minor incidents were detected such as starting sexual relationships with the risk of HIV infection, vagrancy, bouts of alcohol consumption, and aggressivity.”<sup>15</sup>

*Explanation:* Harms involve the possible adverse or negative consequences of an intervention, whereas unintended effects are those not hypothesised to occur as a result of an intervention but may be harmful or beneficial.<sup>78</sup> Though often not examined or thought not even to be possible, social and psychological interventions have the potential to harm or have unintended effects.<sup>79</sup> Though some overlap exists, social and psychological

interventions may have adverse events and unintended consequences that differ from those sought and detected in medical trials. Potential harms that may be unique to social and psychological interventions include increased substance use, crime, or unemployment. When reporting adverse events or unintended outcomes, authors should indicate how these were defined and measured, as well as the frequency of each per trial group.

Ideally, potential harms are identified in the protocol and relate to the theory of how an intervention is hypothesised to work (Item 2b). However, some harms may not be in the protocol either because there was no protocol, the authors did not think to include harms in the protocol, or harms identified in the trial were not expected beforehand. Results from qualitative investigations that reveal unintended consequences of the intervention may be delineated where relevant. For more complex social interventions working on multiple levels, investigators should report adverse consequences for groups or places as well as individuals. Inequity caused by the intervention, such as increased socioeconomic deprivation in a community, is one such example.

### ***3.20 Discussion: Limitations***

*Item 20. Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses*

*Example:* “There are a number of limitations to this study that should be noted. As described previously, statistical conclusion validity in the argumentation analysis was limited by not having a specific argumentation pretest covariate in the model, and instead a knowledge / reasoning pretest value for each student was used because a correlation was expected. Our claims about retention are similarly tempered (no argumentation pretest or no knowledge/reasoning retention measure). Other limitations of this study include the small sample size (58 students) and the short length of the intervention (10 hours of instruction, 4 hours of testing), yet the fact that we found significant and consistent differences despite these limitations speaks to the strength of the effect.... Despite the teacher in this study having many years experience teaching both traditional and inquiry-based materials, he is undoubtedly more of an advocate of an inquiry-based approach. However, we believe the benefits of controlling variables by having the same teacher in both sections outweighed the potential bias created by a teacher being more comfortable in one approach than the other, and findings such as the comparable levels of student engagement shown in Table 4 suggest that the treatments were not strongly teacher-biased.”

*Explanation:* Authors should consider potential biases that may be present in a trial, such as issues related to the precision of effect estimates, heterogeneity in delivery and uptake of the interventions, and the use of multiple analyses. Intervention implementation is a key issue that should always be addressed in the social and psychological intervention literature. When conducted and reported, process evaluation limitations should be discussed, such as whether measurement of intervention fidelity or focus groups to discuss implementation may influence the natural intervention setting. Given the prevalence of self-report measures in social and psychological intervention, issues to do with comprehension of measures (e.g., literacy levels, cultural translation of concepts) may also need to be addressed.

### ***3.21 Discussion: Generalisability***

#### *Item 21. Generalisability (external validity, applicability) of the trial findings*

*Example:* “The study was conducted at a research center of a general hospital. Although the ethnic makeup of the sample roughly approximated that of the county in which our research center was situated (Middlesex County, Massachusetts), which includes 80.3% European Americans, 4.3% African Americans, 8.6% Asian Americans, 5.8% people of Latino origin, and 1.4% people of two or more races, the sample included fewer African Americans, more biracial children, and participants with higher parental education and socioeconomic status than the general Middlesex County community. Further studies are needed to examine the protocol’s efficacy in samples with greater socioeconomic and ethnic diversity, as well its effectiveness in community mental health settings. In addition, our criteria allowed for the exclusion of children judged too uncooperative or distractible to take part in the treatment (two children) or children deemed too clinically severe to wait 6 months to receive treatment, based on severe mood disorder, severe social isolation, severe impairment in school function or attendance, or severe OCD (obsessive compulsive disorder) (a total of seven children). These criteria generally excluded children who clinically would not be administered CBT (cognitive behavioral therapy) for anxiety disorders as their first treatment (i.e., they might be offered such treatment after their other symptoms were addressed). Therefore, study results can be generalized only to children whose anxiety disorders are not so severe as to cause school refusal or severe social isolation. In other regards, however, the sample appeared representative of clinical samples, with high comorbidity of anxiety disorders and with 69% in the borderline or clinical range on the CBCL (child behavior checklist) Internalizing scale. In addition, our extensive intake assessment battery, which required a total of four parent and/or child visits prior to randomization, deterred as many as 1 in 4 potential participants and may have selected for families who were especially motivated to take part in treatment.”<sup>60</sup>

*Explanation:* The generalisability of an RCT is important to understand yet a complicated judgment for authors to make about their own trials. Compared to such judgements, the reader would likely benefit more from details being provided in the methods and results section about the nature of the sample, how the sample was recruited, and important features of the setting and intervention, in order to make their own judgments about the applicability of trial findings to their particular contexts. Nonetheless, policy-makers and practitioners are often interested in the author's judgments about the generalisability of a trial's findings. Formal theories for causal generalisation may be useful to consult.<sup>38</sup>

When making such conclusions in the discussion section, authors should be clear how their statements about the populations and settings to which the results are likely to apply are commensurate with the trial design and execution. Based on theoretical understandings of the intervention at hand, key influences to consider include the restrictiveness of recruitment practices and eligibility criteria, characteristics of the sample, intervention implementation, the choice of the comparator, how and what outcomes were assessed, length of follow-up, and features of the setting.<sup>19</sup> For example, the most defensible claims to generalisability are likely to be made about populations and settings that made up the sampling frame. Authors should be clear about how restrictions in recruitment procedures limit inferences about the applicability of findings to certain members of the target population.<sup>19</sup> Aspects to consider might be the geographic location, the immediate setting, and incentives provided. If an active intervention served as the control, consideration should be given to the services that were delivered. Particularly for effectiveness trials, authors should discuss issues related to the sustainability of the intervention, as well as its potential to be adopted and implemented in non-experimental settings. Relevant facilitators and barriers to implementation could be related to intervention roll-out or organisational capacity; for example, many multi-level public

health interventions take quite some time to set-up.

### ***3.22 Discussion: Interpretation***

*Item 22. Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence*

*Example:* ““The results from this trial differed from previous CBT (cognitive behavioural therapy) trials in two key areas. Only one patient (3%) did not complete the treatment. Previous IBS (irritable bowel syndrome) studies suggest that drop-out rates from CBT can be as high 40%. This may be because traditional CBT requires a substantial time commitment from patients. The most common reasons for dropping out are being unable to take time off work or childcare commitments. Having fewer sessions and sessions on the telephone may make the therapy more widely available. In addition, presenting treatment as self-management of a chronic condition rather than as a psychological therapy may be more acceptable to IBS patients. The treatment effects for symptom severity in this study are larger than those reported in many other CBT trials. This may be because of differences in the patient cohorts. In the London-based primary-care CBT trial, the mean WSAS (Work and Social Adjustment Scale) score at baseline was 15 and the mean HADS (Hospital Anxiety and Depression Scale) depression score was around 7. In the current study, the median WSAS was 5 and the mean HADS depression score was 4. As our study did not rely on GP (general practitioner) referral we may have accessed a cohort that seldom gets offered therapeutic intervention or perhaps even gets diagnosed. This is important, as our results suggest that treatment effects may be greater if patients are less disabled by their symptoms and less depressed. There is certainly evidence that depression in IBS is related to poorer treatment outcome. This study indicates that early intervention and diagnosis may not only make treatment more effective but also prevent the illness becoming more chronic and refractory to treatment.”<sup>20</sup>

*Explanation:* Authors should provide the readers with a brief interpretation of the findings in light of the study objectives, including statements of support for primary and secondary objectives or hypotheses.<sup>19</sup> This interpretation should consider the strengths of the trial as well as limitations, particularly the success of and barriers to implementing the intervention (Item 20).<sup>12</sup> As for the introduction, authors should again contextualise results by discussing relevant systematic reviews or the totality of evidence for other interventions thought to be effective for the targeted problems, but now with the results of this study in mind.<sup>80</sup> Authors should delineate how the trial is congruent or disagrees with the results of other relevant literature. Relevant evidence may include information about evaluations of the same or similar interventions, evaluations of different interventions on the same

population and problem, studies on any proposed intervention mediators or moderators, and pertinent theoretical literature. If significant limitations were experienced, authors may wish to discuss the most plausible explanations for results other than differences in the effectiveness of the experimental and comparator interventions.

The interpretation may also include implications for future research, practice, and policy.<sup>19</sup> Authors should note what future research would be helpful in addressing gaps in current knowledge as well as limitations in the trial. For interventions substantively based on theory (Item 2b), authors should discuss the implications of the trial for theory and evidence on the causal pathways or mechanisms through which the intervention was thought to work.<sup>19</sup> For large trials aiming to address practice or policy questions, or if the trial data are combined with those of a previous meta-analysis, authors may comment on the practical significance of the findings and the implications of findings to practice and policy.<sup>24</sup> Authors should be sure not to amplify the implications of their findings.

### ***3.23 Important Information: Registration***

*Item 23. Registration number and name of trial registry*

*Example:* “This trial is registered at [www.controlled-trials.com](http://www.controlled-trials.com) (ISRCTN97730834).”<sup>51</sup>

*Explanation:* Trial registration serves many purposes, such as providing a publicly available pre-specification of trial design and hypotheses, aiding in systematic review and meta-analysis efforts, supplying a better record of trials being conducted, and working as a barrier to publication bias or selective outcome reporting biases. Ideally, trials should be registered prospectively before baseline data have been collected in order to achieve all of these aims, although registration after the trial has started or even completed can still address some aims (e.g., to update interested readers on other manuscripts or materials about the trial). Authors should note in the manuscript at which point the trial was

registered (e.g., *a priori*, after baseline assessment, after all outcome data were assessed).

At the time of publication, few social and psychological intervention trials report trial registration numbers.<sup>17</sup> Furthermore, many participants in the CONSORT-SPI Delphi process commented that appropriate registries do not exist for social and psychological intervention trials. To the contrary, registries used in clinical medicine—such as [clinicaltrials.gov](http://clinicaltrials.gov) or the WHO trials register—equally apply to social and psychological intervention RCTs, with published instructions on how to use such registries available in the social and behavioural science literature.<sup>81</sup> At a minimum, registries typically have authors list the dates and location of the trial, its objectives and hypotheses, basic information about the interventions being studied, outcome measurement and analysis plans, and ethical approval. All of these apply to social and psychological intervention trials. Furthermore, several registries exist specifically for social and psychological interventions, such as the American Economic Association RCT Registry for social science trials and the International Initiative for Impact Evaluation's (3ie) Registry for International Development Impact Evaluations (RIDIE).

### ***3.24 Important Information: Protocol***

*Item 24. Where the full trial protocol can be accessed, if available*

*Example:* “Details of the study method have been described elsewhere (de Graaf et al., 2008).”

*Explanation:* Ideally, many of the details about trial design would be contained in a publicly accessible protocol that the subsequent trial could reference, freeing space in the trial report to only note differences from the protocol. When existent, these typically are either peer-reviewed published manuscripts, or reports on a website. For further reference, guidance on developing and reporting protocols has recently been published.<sup>82</sup>

### 3.25 Important Information: Funding

#### *Item 25. Sources of funding and other support, role of funders*

*Example:* “This work was supported by a research grant from the National Institute on Drug Abuse (R01 DA015183-05) with cofunding from the National Cancer Institute, the National Institute of Child Health and Human Development, the National Institute of Mental Health, and the Center for Substance Abuse Prevention.”<sup>83</sup>

*Explanation:* Authors should identify and describe relevant sources of funding or other support received while conducting the trial. These sources could have provided monetary or material support. Authors should indicate the name of any person or entity receiving funds or awards, the name of the funder, and the award number to inform the reader. They should also specifically state if these sources of support had any role in the design, conduct, analysis, and reporting of the trial. Donation of significant physical space to conduct the study or of materials such as proprietary outcome measures should also be acknowledged, so long as confidentiality agreements are not in place.<sup>19</sup>

#### *Extended CONSORT-SPI Item 25. Declaration of any other potential interests*

*Example 1:* “D.R.H-B. and A.H. have received honoraria from Reed Medical Education (a company working as a logistics collaborator for the MGH Psychiatry Academy). The education programs conducted by the MGH Psychiatry Academy were supported, in part, through independent medical education grants from pharmaceutical companies, including AstraZeneca, Lilly, McNeil Pediatrics, Janssen, Bristol-Myers Squibb, Shire, Forest Laboratories, Inc., Sanofi Aventis, and Pfizer.”<sup>60</sup>

*Example 2:* “The contributions of GP in reviewing the creative process instrument a few times are gratefully acknowledged. GP authored a few books such as Evidence based teaching—A practical approach (2006), Teaching today: A practical guide (2004) and How to be better at creativity (1996). He works for the Learning Skills Development Agency as a consultant on their Raising Quality and Achievement programme, assisting Action Research Development Projects in colleges, and assisting the Quality Improvement Team. He is a visiting examiner for the Institute of Education at London University. His experience includes physics teaching, managing teacher training, Inclusive Learning Facilitator, being a staff development officer, and managing college lesson observation.”<sup>84</sup>

*Explanation:* Authors should disclose and describe any other relationships that may be

perceived as potential conflicts of interest with regards to the design, conduct, analysis, or reporting the trial. Authors are advised to err on the side of caution, as the discovery of potential conflicts of interest not disclosed in a manuscript are potentially more detrimental than divulging potential interests that are later deemed by editors to be inconsequential and unnecessary for mentioning in publications.<sup>19</sup> Such conflicts may involve economic or commercial benefit from the use of the intervention under study. In comparison to the previous item on funding, this extended item also notes that, in social and psychological intervention literature, potential conflicts are not always commercial or financial. Moreover, not all journals in the social and behavioural sciences have policies in place to require authors to report conflicts of interests, either non-financial or at all. Clear statements about the researchers' relationship to the intervention are critical. Allegiance to or professional training on an intervention are examples,<sup>18</sup> particularly when employment of trial investigators and staff, or the sustainability of organisations involved in implementation, rests on the credibility of the intervention and the trial results.<sup>85</sup>

### ***3.26 Important Information: Stakeholder Involvement***

*New CONSORT-SPI Item. Any involvement of the intervention developer in the design, conduct, analysis, and reporting of the trial*

*Example:* “The designers of the system have worked since the inception of the project to ensure that the worked examples and self-explanation prompts are aligned with the course content and that the language used parallels that of the lectures and textbook.”<sup>86</sup>

*Explanation:* Intervention developers often are involved in RCTs evaluating their interventions, either as part of the research team or as some form of advisor. The intervention developer could be an individual or team of individuals, or it could be a purveyor organisation involved in the continual development and upkeep of an intervention that might benefit from positive results of the trial. Involvement of the

intervention developer could potentially lead to bias, as reputation and career advancement could be linked to the success or failure of the trial. However, it also could serve as a quality control by increasing the potential of delivering interventions as designed. The authors need to transparently report such information for the reader to appraise and come to their own conclusions.

The relationship between the authors and the intervention under investigation needs to be known. These relationships are not always direct or formal. For example, those conducting the study may work in the same organisation as or be a former student of the developer. The researchers may be on an advisory or consultancy board related to the intervention. The researchers may have other types of allegiance to the intervention or an alliance with its underlying ideology, for example when they were trained in the intervention and provide it in professional practice outside of the research environment. Those evaluating “brand name” programmes should also note any patents, intellectual rights, copyrights, or proprietary rights they may have for the intervention.

*New CONSORT-SPI Item. Other stakeholder involvement in trial design, conduct, and/or analyses*

*Example:* “The SKCHH [Seattle–King County Healthy Homes] project was designed as a community-based participatory research project with overall sponsorship by Seattle Partners for Healthy Communities, an Urban Research Center funded by the U.S. Centers for Disease Control and Prevention. Seattle Partners is a multidisciplinary partnership of community agencies, community activists, public health professionals, academics, and health providers that supports community-based participatory research addressing social determinants of health.... Both the Seattle Partners Board and the steering committee sought to assure that the project benefited all participants. This led to the staggered intervention design with low- and high-intensity groups. This design assured that low-intensity group participants initially received some immediate benefit (including interventions known to be useful, such as bedding encasements) while ultimately receiving all the benefits accorded the high-intensity group. While this design may have reduced the study’s power to demonstrate an effect of the high-intensity intervention relative to a “pure” control group receiving no intervention, we felt such a design was not ethical.”<sup>87</sup>

*Explanation:* Researchers are increasingly called to consult or collaborate with those who have a direct stake in the outcomes of the trial, such as those who would be delivering or receiving the intervention once the trial is completed. Such methods help those involved with research as well as those being intervened on have their voices heard. Stakeholder involvement can also help to better ensure the acceptability, implementability, and sustainability of interventions as they move from research to real-world settings.

Authors should note which stakeholders were involved in the trial and how they were involved. Involvement may include community-based needs assessments, choice of intervention content or adaptation, choice of outcomes and their measurement, analysis of trial data, and developing manuscripts about the trial. An increasingly common practice is to develop Community Advisory Boards (CABs) that assist with these matters. If no stakeholders from the target population were involved, the authors should make this clear.

*New CONSORT-SPI Item. Incentives offered as part of the trial*

*Example:* “Participants were informed that they would receive \$10 for the screening survey, \$25 for the baseline survey, and \$30 for a follow-up survey. They were to receive a \$5 bonus for completion of any survey within 48 hours. Students also were informed about attending a 1-hour education program and brief evaluation, for which they would receive \$10.... The parents were sent a letter explaining the study, a consent form, and a \$10 check and were asked to complete a survey assessing parent-teen communication. The letter informed parents that they would be sent materials to read and evaluate sometime within the next year, for which they would receive \$15.”<sup>88</sup>

*Explanation:* Incentives offered to participants, providers, and organisations involved in an RCT can influence trial attrition, engagement with the intervention, and quality of delivery. Incentives are typically offered for completing screening, for entering the trial, for engaging with the intervention, and for completing outcome assessments. It is important to report the nature of incentives, such as monetary compensation, free meals, free transport, free access to an otherwise costly service, extra credit for a school course, or even coercion

(i.e., the incentive is to not experience an aversive state).<sup>18</sup> Authors should also note whether incentives differ by trial group, such as compensation for participants attending the intervention in the experimental group but not in the control group. Incentives are important to know when considering replication or roll-out, particularly those related to using the intervention (with regards to its effectiveness) and recruitment (with regards to real-world referrals and uptake). Other retention efforts (e.g., several phone calls to participants to obtain outcome data) may be reported when applicable.

#### **4. Discussion**

This CONSORT-SPI E&E document is intended to assist authors in writing reports of social and psychological intervention RCTs in accordance with the CONSORT-SPI checklist, serving as a “users manual”. Such guidance could even help social and behavioural scientists with trial design, as the goal-posts for the ultimate report are known in advance, and references to appropriate methodological literature are provided above. The CONSORT-SPI E&E can also assist editors, peer-reviewers, and research funders better understand what is essential for high-quality social and psychological intervention trials for their assessments of submitted publications and grant proposals. Assessments facilitated by the CONSORT-SPI E&E hopefully will lead to higher quality trials from the design stage to clear and comprehensive reports at the publication stage.

##### ***4.1 Strengths and Limitations***

The examples of proper reporting, as well as the rationale, are at a general level for the above draft, and may not equally apply to all disciplines being targeted by this guideline. Some of the optional detail suggested in this E&E may not be possible for all trials (e.g., place-based intervention trials) or may require a word count larger than the

limits of some journals (e.g., when a trial has more than two arms). Furthermore, many examples are drawn from the psychological literature, as they tended to have the best quality reports amongst trials in the systematic literature review in Chapter 3. As the CONSORT-SPI project continues, versions of this E&E statement that are more attuned to the specific theories, intervention techniques, outcomes, and contexts of particular disciplines should better ensure proper use of the CONSORT-SPI checklist within those disciplines.

Judging adherence to some items is more subjective than for others—particularly those related to settings and interventions. Standards related to context are limited by the weaknesses in this field in general of conceptualising theories and context. Much work still needs to be done to be able to operationalise which dimensions of context really are crucial, how they are important, how to measure them, and how to analyse their influence. Hopefully better reporting of setting and the target population within trials will help researchers unpack the nitty-gritty of the dimensions that require detailed reporting. While guidance is provided on how to adhere to items, adherence can differ for a variety of reasons—e.g., maturity of the intervention, complexity of the intervention. Ultimately, it is up to the authors, editors, and peer-reviewers to decide which details are important to include. Guidelines do not take away the expertise needed to write good trial reports, but serve as a cognitive reminder as to what authors need to report.<sup>89</sup> However, those using the checklist may not read the E&E thoroughly, if at all. There will be explicit, visible instruction in the CONSORT-SPI Extension Statement for readers to use the E&E along with the checklist.

#### ***4.2 Implications***

The development of the CONSORT-SPI E&E is on-going. This draft template will

be refined by the Project Executive to serve as the basis for the different versions of the E&E document tailored to specific disciplines. Namely, this draft E&E template is currently planned to be disseminated to the CONSORT-SPI Project Executive for editing until a final version is agreed. The template will then be disseminated to the writing groups of the various E&E statements for CONSORT-SPI, who will address issues like wording, reporting examples provided, and best ways to explain the rationale for items for stakeholders in their area. The final versions of these E&E documents will include theory and evidence from the wider social and psychological intervention research literature for supporting information as needed.

## **5. Conclusion**

As with the CONSORT 2010 checklist and E&E, the CONSORT-SPI E&E will be recommended for use alongside the CONSORT-SPI checklist. The CONSORT-SPI Extension is intended to join the wider “CONSORT family” in its attempt to improve the quality of trial reports. Developed according to best practices—i.e., systematic reviews and rigorous consensus development process—the guidance should provide the most appropriate standards and advice for social and psychological intervention research. Moreover, as many areas of social and psychological intervention research are currently under rapid development, new evidence as well as feedback will be used to guide future updates of this guidance. However, simply developing a guideline does not guarantee its use. A coordinated dissemination and implementation plan is needed to detail how potential users are expected to become aware of new guidance as well as use it. The next chapter will turn to this issue.

## References

1. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
2. Moher D, Weeks L, Ocampo M, et al. Describing reporting guidelines for health research: A systematic review. *Journal of Clinical Epidemiology*. 2011;64(7):718-742.
3. Chan A, Tetzlaff JM, Gøtzsche PC, et al. SPIRIT 2013 explanation and elaboration: Guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.
4. Moher D, Hopewell S, Schultz KF, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
5. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*. Oct 2009;62(10):e1-34.
6. Campbell Collaboration. <http://campbellcollaboration.org/>. 2013.
7. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *The Lancet*. 2008;371:281-283.
8. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: Extension to cluster randomised trials. *BMJ*. 2012;345:e5661.
9. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
10. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT Group.

- Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: Explanation and Elaboration. *Annals of Internal Medicine*. 2008;148:295-309.
11. Flay BR, Biglan A, Boruch RF, et al. Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*. 2005;6:151-175.
  12. Des Jarlais DC, Lyles C, Crepaz N, the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*. 2004;94(3):361-366.
  13. Moore G, Audrey S, Barker M, et al. Process evaluation in complex public health intervention studies: The need for guidance. *Journal of Epidemiology & Community Health*. 2013;68(2):101-102.
  14. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698-702.
  15. Girón M, Fernández-Yañez A, Mañá-Alvarenga S, Molina-Habas A, Nolasco A, Gómez-Beneyto M. Efficacy and effectiveness of individual family intervention on social and clinical functioning and family burden in severe schizophrenia: A 2-year randomized controlled study. *Psychological Medicine*. 2010;40:73-84.
  16. Petticrew M, Roberts H. *Systematic reviews in the social sciences: A practical guide*. Oxford, UK: Blackwell Publishing Ltd; 2006.
  17. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.
  18. Davidson KW, Goldstein M, Kaplan RM, et al. Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine*.

- 2003;26(3):161-171.
19. Cooper H. *Reporting research in psychology: How to meet journal article reporting standards*. Washington, DC: APA; 2011.
  20. Moss-Morris R, McAlpine L, Didsbury LP, Spence MJ. A randomized controlled trial of a cognitive behavioural therapy-based self-management intervention for irritable bowel syndrome in primary care. *Psychological Medicine*. 2010;40:85–94.
  21. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: Explanation and Elaboration. *PLoS Medicine*. 2008;5(1):e20.
  22. Anderson T, Guajardo JF, Luthra R, Edwards KM. Effects of clinician-assisted emotional disclosure for sexual assault survivors: A pilot study. *Journal of Interpersonal Violence*. 2010;25:1113-1131.
  23. American Educational Research Association. Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*. 2006;35(6):33-40.
  24. Moberg-Mogren E, Nelson DL. Evaluating the quality of reporting occupational therapy randomized controlled trials by expanding the CONSORT criteria. *The American Journal of Occupational Therapy*. 2006;60:226-235.
  25. McKirnan DJ, Tolou-Shams M, Courtenay-Quirk C. The Treatment Advocacy Program: A randomized controlled trial of a peer-led safer sex intervention for HIV-infected men who have sex with men. *Journal of Consulting and Clinical Psychology*. 2010;78(6):952–963.
  26. Corbalan G, Paas F, Cuypers H. Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers & Education*. 2010;55:692-703.

27. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychology*. 2007;27:379-387.
28. Webb MS, de Ybarra DR, Baker EA, Reis IM, Carey MP. Cognitive-behavioral therapy to promote smoking cessation among African American smokers: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*. 2010;78(1):24-33.
29. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW, for the CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT Statement. *JAMA*. 2006;294(10):1152-1160.
30. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ*. 2008 337:a2390.
31. Barrio C, Yamade AM. Culturally based intervention development: The case of Latino families dealing with schizophrenia. *Research on Social Work Practice*. 2010;20(5):483-492.
32. Shamseer L, Sampson M, Bukutu C, et al. CONSORT extension for N-of-1 trials (CENT) guidelines. *BMC Complementary and Alternative Medicine*. 2012;12(Suppl 1):P140.
33. Cunha DB, Souza BdsNd, Pereira RA, Sichieri R. Effectiveness of a randomized school-based intervention involving families and teachers to prevent excessive weight gain among adolescents in Brazil. *PLoS One*. 2013;8(2):e57498.
34. Chang HY, Quintana C, Krajcik JS. The impact of designing and evaluating molecular animations on how well middle school students understand the particulate nature of matter. *Science Education*. 2010;94(1):73-94.
35. Lindsey MA, Hayward A, DePanfilis D. Gender differences in behavioral outcomes among children at risk of neglect: Findings from a family-focused

- prevention intervention. *Research on Social Work Practice*. 2010;20(6):572-581.
36. Higginson A, Mazerolle L. Reporting guidelines for social and psychological intervention trials: Considering the special context of crime and justice interventions. *Journal of Experimental Criminology*. 2013: Online first. DOI: 10.1007/s11292-11013-19186-z.
37. Buysse V, Castro DC, Peisner-Feinberg E. Effects of a professional development program on classroom practices and outcomes for Latino dual language learners. *Early Childhood Research Quarterly*. 2010;25:194-207.
38. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth; 2002.
39. McKay JR, Van Horn DHA, Oslin DW, et al. A randomized trial of extended telephone-based continuing care for alcohol dependence: Within-treatment substance use outcomes. *Journal of Consulting and Clinical Psychology*. 2010;78(6):912-923.
40. Hoffmann TC, Glasziou P, Boutron I, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687.
41. Mohler R, Bartoszek G, Kopke S, Meyer G. Proposed criteria for reporting the development and evaluation of complex interventions in healthcare (CReDECI): Guideline development. *International Journal of Nursing Studies*. 2012;49:40-46.
42. Michie S, Richardson M, Johnston M, et al. The Behavior Change Technique Taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*. 2013;46:81-95.
43. Perera R, Heneghan C, Yudkin P. A graphical method for depicting randomised

- trials of complex interventions. *BMJ*. 2007;334:127-129.
44. Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F. *Implementation research: A synthesis of the literature*. Tampa, FL: National Implementation Research Network, University of South Florida;2005.
  45. Boivin MJ, Busman RA, Parikh SM, et al. A pilot study of the neuropsychological benefits of computerized cognitive rehabilitation in Ugandan children with HIV. *Neuropsychology*. 2010;24(5):667–673.
  46. Torgerson DJ. Contamination in trials: Is cluster randomisation the answer? *BMJ*. 2001;322:355-357.
  47. Alvarez KM, Donohue B, Carpenter A, Romero V, Allen DN, Cross C. Development and preliminary evaluation of a training method to assist professionals in reporting suspected child maltreatment. *Child Maltreatment*. 2010;15(3):211-218.
  48. Twohig MP, Hayes SC, Plumb JC, et al. A randomized clinical trial of acceptance and commitment therapy versus progressive relaxation training for obsessive-compulsive disorder. *Journal of Consulting and Clinical Psychology*. 2010;78(5):705-716.
  49. West R. Providing full manuals and intervention descriptions: Addiction policy. *Addiction*. 2008 103:1411.
  50. Stice E, Rohde P, Gau JM, Wade E. Efficacy trial of a brief cognitive-behavioral depression prevention program for high-risk adolescents: Effects at 1- and 2-year follow-up. *Journal of Consulting and Clinical Psychology*. 2010;78(6):856-867.
  51. de Bruin M, Hospers HJ, van Breukelen GJ, Kok G, Koevoets WM, Prins JM. Electronic monitoring-based counseling to enhance adherence among HIV-infected patients: A randomized controlled trial. *Health Psychology*. 2010;29(4):421-428.

52. Australian National Statistical Service.  
<http://www.nss.gov.au/nss/home.nsf/pages/Sample+size+calculator>. Accessed on 15 June 2014.
53. Tracy SK, Hartz DL, Tracy MB, et al. Caseload midwifery care versus standard maternity care for women of any risk: M@NGO, a randomised controlled trial. *Lancet*. 2013;382(9906):1723-1732.
54. Chalder T, Deary V, Husain K, Walwyn R. Family-focused cognitive behaviour therapy versus psycho-education for chronic fatigue syndrome in 11- to 18-year-olds: A randomized controlled treatment trial. *Psychological Medicine*. 2010;40:1269–1279.
55. Gleeson JFM, Cotton SM, Alvarez-Jimenez M, et al. Family outcomes from a randomized control trial of relapse prevention therapy in first-episode psychosis. *Journal of Clinical Psychiatry*. 2010;71(4):475-483.
56. Carey MP, Senn TE, Venable PA, Coury-Doniger P, Urban MA. Brief and intensive behavioral interventions to promote sexual risk reduction among STD clinic patients: Results from a randomized controlled trial. *AIDS & Behavior*. 2010;14:504-517.
57. Mueser KT, Pratt SI, Bartels SJ, et al. Randomized trial of social rehabilitation and integrated health care for older people with severe mental illness. *Journal of Consulting and Clinical Psychology*. 2010;78(4):561-573.
58. McKay JR, Lynch KG, Coviello D, et al. Randomized trial of continuing care enhancements for cocaine-dependent patients following initial engagement. *Journal of Consulting and Clinical Psychology*. 2010;78(1):111-120.
59. Witkiewitz K, Bowen S. Depression, craving, and substance use following a randomized trial of mindfulness-based relapse prevention. *Journal of Consulting*

- and Clinical Psychology*. 2010;78(3):362-374.
60. Hirshfeld-Becker DR, Masek B, Henin A, et al. Cognitive behavioral therapy for 4- to 7-year-old children with anxiety disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*. 2010;78(4):498-510.
  61. Wilson DB. Comment on “Developing a reporting guideline for social and psychological intervention trials”. *Journal of Experimental Criminology*. 2013;9:375-377.
  62. Montgomery GH, Hallquist MN, Schnur JB, David D, Silverstein JH, Bovbjerg DH. Mediators of a brief hypnosis intervention to control side effects in breast surgery patients: Response expectancies and emotional distress. *Journal of Consulting and Clinical Psychology*. 2010;78(1):80-88.
  63. MacPherson L, Tull MT, Matusiewicz AK, et al. Randomized controlled trial of behavioral activation smoking cessation treatment for smokers with elevated depressive symptoms. *Journal of Consulting and Clinical Psychology*. 2010;78(1):55-61.
  64. Little RJ, Rubin DB. *Statistical analysis with missing data*. 2nd ed. Hoboken, NJ: Wiley; 2002.
  65. Rodriguez ML, Dumont K, Mitchell-Herzfeld SD, Walden NJ, Greene R. Effects of Healthy Families New York on the promotion of maternal parenting competencies and the prevention of harsh parenting. *Child Abuse & Neglect*. 2010;34(10):711-723.
  66. Gardner F, Hutchings J, Bywater T, Whitaker C. Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *Journal of Clinical Child & Adolescent Psychology*. 2010;39(4):568-580.

67. Cinciripini PM, Blalock JA, Minnix JA, et al. Effects of an intensive depression-focused intervention for smoking cessation in pregnancy. *Journal of Consulting and Clinical Psychology*. 2010;78(1):44-54.
68. Henderson CE, Dakof GA, Greenbaum PE, Liddle HA. Effectiveness of multidimensional family therapy with higher severity substance-abusing adolescents: Report from two randomized controlled trials. *Journal of Consulting and Clinical Psychology*. 2010;78(6):885-897.
69. Kaner E, Bland M, Cassidy P, et al. Effectiveness of screening and brief alcohol intervention in primary care (SIPS trial): Pragmatic cluster randomised controlled trial. *BMJ*. 2013;346:e8501.
70. Wood MD, Fairlie AM, Fernandez AC, et al. Brief motivational and parent interventions for college students: A randomized factorial study. *Journal of Consulting and Clinical Psychology*. 2010;78(3):349-361.
71. Hodnett RH, Faulk K, Dellinger A, Maher E. *Evaluation of the statewide implementation of a parent education program in Louisiana's child welfare agency: The Nurturing Parenting Program for Infants, Toddlers, and Pre-School Children*. Seattle: Casey Family Programs; 2009.
72. Mills M, Loney P, Jamieson E, et al. A primary care cardiovascular risk reduction clinic in Canada was more effective and no more expensive than usual on-demand primary care: A randomised controlled trial. *Health & Social Care in the Community*. 2010;18(1):30-40.
73. Manne S, Jacobsen PB, Ming ME, Winkel G, Dessureault S, Lessin SR. Tailored versus generic interventions for skin cancer risk reduction for family members of melanoma patients. *Health Psychology*. 2010;29(6):583-593.
74. Welch V, Petticrew M, Tugwell P, et al. PRISMA-Equity 2012 extension:

- Reporting guidelines for systematic reviews with a focus on health equity. *PLoS Medicine*. 2012;9(10):e1001333.
75. Priebe S, Yeeles K, Bremner S, et al. Effectiveness of financial incentives to improve adherence to maintenance treatment with antipsychotics: Cluster randomised controlled trial. *BMJ*. 2013;347:f5847.
76. Driessen E, Van HL, Don FJ, et al. The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: A randomized clinical trial. *American Journal of Psychiatry*. 2013;170:1041-1050.
77. Lipsey MW, Wilson DB. *Practical meta-analysis*. Thousand Oaks, CA: Sage; 2001.
78. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Annals of Internal Medicine*. 2004;141:781-788.
79. McCord J. Cures that harm: Unanticipated outcomes of crime prevention programs. *The ANNALS of the American Academy of Political and Social Science*. 2003;587:16-30.
80. Chalmers I. Trying to do more good than harm in policy and practice: The role of rigorous, transparent, up-to-date evaluations. *The ANNALS of the American Academy of Political and Social Science*. 2003;589(22):22-40.
81. Harrison BA, Mayo-Wilson E. Trial registration: Understanding and preventing reporting bias in social work research. *Research on Social Work Practice*. 2014;24(3):372-376.
82. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: Defining standard protocol items for clinical trials. *Annals of Internal Medicine*. 2013;158(3):200-207.

83. Cho K, MacArthur C. Student revision with peer and expert reviewing. *Learning & Instruction*. 2010;20:328-338.
84. Leng EY, Ali WZbW, Mahmud Rb, Baki R. Computer games development experience and appreciative learning approach for creative process enhancement. *Computers & Education*. 2010;55:1131-1144.
85. Eisner M. *The South Carolina Triple P System population trial to prevent child maltreatment: Seven reasons to be sceptical about the study results*. University of Cambridge Institute of Criminology: Violence Research Centre Working Paper; 2014.
86. Biesinger K, Crippen K. The effects of feedback protocol on self-regulated learning in a web-based worked example learning environment. *Computers & Education*. 2010;55:1470-1482.
87. Krieger J, Takaro TK, Allen C, et al. The Seattle–King County Healthy Homes Project: Implementation of a comprehensive approach to improving indoor environmental quality for low-income children with asthma. *Environmental Health Perspectives*. 2002;110(Supplement 2):311-322.
88. Turrisi R, Larimer ME, Mallett KA, et al. A randomized clinical trial evaluating a combined alcohol intervention for high-risk college students. *Journal of Studies on Alcohol and Drugs*. 2009;70(4):555-567.
89. Gawande A. *The checklist manifesto: How to get things right*. London: Profile Books; 2010.

**Chapter 7:**  
**A Conceptual Framework for Disseminating and Implementing Reporting  
Guidelines: An Example using CONSORT-SPI**

**Abstract**

**Background:** The CONSORT-SPI guidance has been developed to assist in the reporting of randomised controlled trials (RCTs) of social and psychological interventions. However, simply creating reporting guidelines has not changed the quality of reporting in health research, as actual guideline use is based on many factors beyond the existence of the guideline. Consequently, many reporting guideline developers have identified the need to improve their dissemination and implementation efforts, such as obtaining endorsement of reporting guidelines by relevant academic journals. The implementation science and clinical guideline literatures offer many valuable lessons that can be adapted for knowledge translation and implementation within the area of reporting guidelines, but to date the links between these two areas have not been explicitly drawn out.

**Methods:** The initial CONSORT-SPI dissemination and implementation strategy was organised into a theory-informed, coordinated approach using a four-step, systematic method. Frameworks and strategies for implementing clinical guidelines and empirically supported interventions were combined with feedback from the CONSORT-SPI consensus meeting as well as existing literature on disseminating reporting guidelines to facilitate this approach. First, those targeted by the strategy and the behaviours that they need to change were identified. Next, barriers to and enablers of behaviour change were specified using theoretical frameworks—the Guideline Implementability Framework, Promoting Action on Research in Health Services Framework, Consolidated Framework for Implementation

Research, and the Theoretical Domains Framework. Informed by the Behaviour Change Wheel, proposed components of the dissemination and implementation strategy to overcome the modifiable barriers and enhance the modifiable enablers were then expounded. Lastly, a plan to measure and understand whether behaviour change occurs was formulated.

**Results:** CONSORT-SPI primarily aims to help authors improve their reporting of social and psychological intervention RCTs, and peer-reviewers improve their assessments of these reports for journals and other publication outlets. Secondary objectives target those behaviours linked to such improvement, including implementation and enforcement of CONSORT-SPI by organisations such as journals and funding agencies, and a supportive environment of reporting guidelines and standards within the social and behavioural sciences. Modifiable barriers and enablers of guideline use were identified across guideline-, individual-, organisational-, and system-levels. Several strategies are proposed for disseminating and implementing CONSORT-SPI according to these barriers and enablers, as well as potential measures to assess and understand changes in reporting social and psychological interventions.

**Conclusions:** Theoretical frameworks helped to identify many possible targets that need to be addressed for better use and implementation of CONSORT-SPI as well as greater support of reporting standards in scientific communities. Guideline developers and organisations dedicated to high-quality reporting of research—such as the EQUATOR Network—may wish to consider solutions informed by behaviour change tools moving forward when coordinating strategies for reporting guideline implementation.

## 1. Introduction

Reporting guidelines have been developed for nearly 20 years now,<sup>1</sup> with the intention of promoting accurate, comprehensive, and transparent reporting by all researchers employing a particular type of study. Over time, best-practices for developing checklists, flow diagrams, and related guidance have become established.<sup>2,3</sup> However, while there is evidence to suggest that the quality of reporting guidelines and their impact on reporting quality have improved in this timeframe,<sup>4,5</sup> much more needs to be done to improve the quality of reporting health and related research.<sup>6</sup>

Simply creating and publishing guidelines—even those of the highest quality—has not solved research reporting problems, partly because actual guideline use is based on many factors beyond the existence of a guideline. In light of continued poor reporting, many reporting guideline developers have identified the need to improve their dissemination and implementation efforts. For example, one strategy with promising empirical support is the clear endorsement of and instructions for adhering to reporting guidelines by academic journals.<sup>7</sup> However, additional solutions to improving the use of reporting guidelines and ultimately the reporting quality of research are likely needed.

The clinical guideline and implementation science literatures provide valuable insights that could potentially improve knowledge translation strategies for reporting guidelines. Implementation of reporting guidelines remains variable and often lacks a theoretical foundation, despite the numerous individual, organisational, and system-level factors influencing the quality of research reporting that are likely to exist. Theoretical frameworks for implementation are progressively being applied in the dissemination of clinical guidelines and empirically supported interventions.<sup>8-11</sup> However, no paper to date has explicitly used these frameworks for reporting guidelines to create a coordinated dissemination and implementation strategy.

It is plausible that multifaceted implementation interventions built upon a careful assessment of barriers and facilitators to reporting guideline use, as well as a coherent theoretical base, may be more effective than single interventions.<sup>11</sup> A dissemination and implementation plan for a reporting guideline like CONSORT-SPI is itself an intervention that needs to address multiple targets and involve multiple stages or phases. In addition, the strategies, activities, and techniques involved in dissemination and implementation plans need to be tailored for different audiences, such as stakeholders with different professional roles and coming from different disciplinary backgrounds.

Effective engagement of various stakeholders has already been instrumental in the CONSORT-SPI project plan from the outset, partly due to the Project Executive's belief that implementation of the guideline would require "champions" from targeted stakeholder groups. These champions could then act as strongly facilitative actors once the guideline is complete, assisting the Project Executive in getting relevant individuals and organisations to endorse and utilise CONSORT-SPI. However, the initial dissemination and implementation strategy for CONSORT-SPI has been atheoretical. A theory-based strategy would likely prove useful for systematically applying existing knowledge about dissemination and implementation of guidelines, as well as current plans for CONSORT-SPI, to a more coordinated approach moving forward. Consequently, this chapter draws on reporting guideline, clinical guideline, and implementation science research literature to develop a dissemination and implementation plan explicitly in terms of actors, barriers and enablers to guideline use, techniques to promote changes in reporting behaviours, and measures to assess the impact of the guideline.

## **2. Methods**

The objective of this chapter is to employ a systematic approach to the development

of a coordinated dissemination and implementation strategy for CONSORT-SPI, which could potentially be applied to other reporting guidelines and be of use to the reporting guideline field as a whole. This action planning<sup>12</sup> follows a recently described, four-step method that uses guiding questions to choose the most appropriate behaviour change techniques when developing theory-informed implementation interventions.<sup>10</sup> The materials used will first be described: i.e., existing guidance for disseminating reporting guidelines from the EQUATOR Network, the original CONSORT-SPI dissemination and implementation strategy, and feedback from the CONSORT-SPI consensus meeting. Next, the four-step process along with theoretical frameworks from the implementation science and clinical guidelines literature used will be explained.

### ***2.1 Materials: Guidance for Disseminating Reporting Guidelines***

The guidance for developing health research reporting guidelines that underpins this project's methodology delineates three post-meeting activities and seven post-publication activities that are relevant to guideline dissemination and implementation.<sup>3</sup> These activities generally involve publication issues (such as open access and copyright), website development, and guideline endorsement and adherence.

Following a consensus meeting, reporting guideline developers need to first write the guidance statement as well as pilot test its checklist and flow diagram, incorporating feedback from piloting into checklist revisions. Developers should also create an E&E document to serve as a users manual for the guidance. As these two documents are being written, developers should also delineate a publication strategy, ideally involving multiple and simultaneous publication of guidance documents across key journals in the targeted area to enhance uptake and dissemination. As this process can be labour-intensive and exhausting, experts in reporting guideline development recommend having one

corresponding author to coordinate publication across various journals. This person would elicit each prospective journal's interest in publishing the guideline, have them agree to a common text, facilitate a common set of peer-reviewers, and identifying one journal to lead on copy-editing.<sup>3</sup> If demand and resources permit, the developers should actively be involved in translation of the guideline into different languages, using robust methods such as back translation to ensure that all translations are appropriate.<sup>3</sup>

Once the guidance is published, developers of the guideline should set up procedures for seeking and addressing feedback and criticisms from all relevant stakeholders, which is particularly useful for any future updates. Developers should also approach journal editors to seek endorsement of the guideline by their journals, especially those involved in its development. Strong, clear wording in the "Instructions to Authors" section of a journal's website should be used, given evidence suggesting that vague instructions diminish the impact of reporting guidance on reporting quality.<sup>7,13-15</sup> In addition to seeking endorsement of the guideline, developers should also promote actual adherence to the checklist, such as requiring authors and peer-reviewers to submit a completed checklist with submissions and reviews.

The team developing the guideline should also set out to evaluate the impact of the guideline on the reporting quality in the targeted area, as well as encourage independent research teams to carry out such evaluations. A recent review of the impact of CONSORT on reporting of RCTs in medical journals indicates the need for more rigorous, prospective studies on the effect of reporting guidelines, dependent on the extent of editorial policies to ensure guideline adherence.<sup>6</sup> Developers should also create and host a dedicated website that would contain the guideline and all relevant materials, lists of those endorsing the guideline, and new information related to the guideline as it emerges. Ideally, the website would be live before publication so that the URL address could be included in all

published materials.<sup>3</sup> Based on feedback and new evidence from relevant scientific literature, an executive group for the reporting guideline should facilitate updates to the checklist and any E&E documents.

## ***2.2 Materials: Original CONSORT-SPI Dissemination and Implementation Strategy***

The systematic review from Chapter 3 identified that stakeholders in the social and behavioural sciences have not, to date, been a large part of the development and dissemination of reporting guidelines related to social and psychological intervention trials, such as previous CONSORT guidelines.<sup>16</sup> From the outset, then, the project has entailed active consultation of and collaboration with stakeholders in order to maximise awareness, understanding, and use of the CONSORT-SPI Extension when reporting social and psychological intervention trials (see Box 1). The desire has been to make sure that the guideline—in addition to being of high quality—is widely read, well-understood by those who are conducting and publishing trials, and useful to those who use the results of such research. To help achieve this end, the dissemination and implementation plan thus far has involved activities that span across all project phases, using several widely-utilised dissemination techniques from previous reporting guideline projects. For example, the project has purposively involved consultation (i.e., the launch paper and the Delphi process) and collaboration (i.e., the International Advisory Group and the consensus meeting) with hundreds of stakeholders internationally.

The finalisation of the CONSORT-SPI checklist, Extension Statement, and E&E will also involve significant engagement with stakeholders. Previous research has identified the need to evaluate the “validity” of consensus methods, yet this is rarely done.<sup>17</sup> Consequently, a final version of the checklist will be distributed to Delphi participants to seek feedback on its wording and content, and to assess whether the consensus meeting

**Box 1. Initial Pathways to Impact Plan for CONSORT-SPI**

Stakeholder	Group	Where	Impact Activity	When	Who
Trialists	Education	1 Major National Conference (UK)  1 International Conference	(1) Campbell Collaboration Conference (2) Publication of guidelines (3) Key education conferences internationally (4) Webinars and training events (5) Policy Brief for Dept of Education (UK and key international corollaries)	(1/2/3) Guideline docs (24 months) (4/5) Training and webinar docs (24 months)	(1/3/4) Steering Committee, Education (2/5) Executive
	Psychology	1 Major National Conference (UK)  1 International Conference	(1) APA; BPS conferences (2) Publication of guidelines (3) Consultative/debriefing meetings with professional groups internationally (4) Webinars and training events (5) Policy Brief for gov. Depts of Health	(1/2) Guideline docs (24 months) (3/4/5) Training and webinar docs (24 months)	(1/3/4) Steering Committee, Psychology (2/5) Executive
	Criminology	1 Major National Conference (UK)  1 International Conference	(1) National Police Improvement Agency Conf. (2) Publication of guidelines (3) Key Criminology conference internationally (4) Webinars and training events (5) Policy Brief for gov. Ministries of Justice	(1/2) Guideline docs (24 months) (3/4/5) Training and webinar docs (24 months)	(1/3/4) Steering Committee, Criminology (2/5) Executive
	Social Work	1 Major National Conference (UK)  1 International Conference	(1) BASW; SSWR Conferences (2) Publication of guidelines (3) Consultative/debriefing meetings with professional groups internationally (4) Webinars and training events (5) Policy Brief for gov Depts of Social Care	(1/2) Guideline docs (24 months) (3/4/5) Training and webinar docs (24 months)	(1/3/4) Steering Committee, Social Work (2/5) Executive
Systematic Reviewers, Guideline Developers	Cochrane	Annual Colloquium	Consultation and adoption	Oct. 2014 and 2015	Executive
	Campbell	Annual Colloquium	Consultation and adoption	May 2015 and 2016	Executive
	Other Producers	WHO, AHRQ SCIE, NICE, NCCMH, EPPI	(1) Conferences and meetings (2) Policy brief, weblinks, and training	(1/2) Guideline docs (24 months)	Executive
Funding Bodies	Various	ESRC, MRC NIH, Other International	(1) Presentations and meetings (2) Web materials and briefs to assist referees	(1/2) Guideline docs (24 months)	Executive
Media + Public	University PR Office	(1) Press Releases (2) Web materials	(1) CONSORT and EQUATOR Websites (2) Brief communications, social media	(1/2) Guideline docs (24 months)	Executive

accurately captured their opinions. Drafts of all guidance documents will also be circulated to consensus group participants to check that the documents (a) accurately represent the decisions made during the meeting, (b) provide examples of good reporting for specific items, and (c) are useful for their intended purpose.<sup>18</sup> Feedback will be incorporated into a CONSORT-SPI Extension Statement that reports the project rationale, process methodology, and final included checklist items.

Moving forward, the dissemination and implementation strategy is intended to be “multi-pronged” in order to ensure that the various stakeholders benefit from this guideline, and to actually have stakeholders (such as the consensus meeting participants) assist with the strategy. Firstly, upon finalising the guideline documents, simultaneous publications in multiple, high-impact, peer-reviewed journals will begin the process of dissemination and uptake.<sup>2</sup> Simultaneous publication of the guidance in multiple journals should increase exposure to the guidance amongst the targeted disciplines. The IAG and consensus meeting participants will help to identify and approach the most appropriate journals in key disciplines to publish the guideline and provide an editorial supporting the guideline. The Project Executive will liaise with journals to agree on a common text for the guideline, as with previous CONSORT guidelines.<sup>19</sup> As mentioned in the previous chapter, different versions of the E&E using examples from particular disciplines are also proposed.

Journals and funding organisations will also be asked to formally endorse the guideline and enforce adherence to it by authors and peer-reviewers. The IAG will also help to identify, contact, and ask editors and representatives from all relevant journals and funding organisations to endorse the guideline. Endorsement is intended to involve clear directions in each journal’s “Instructions to Authors” or organisation’s submission policies that the guideline should be followed in all social and psychological intervention RCT reports. Ideally, these policies will also ask peer-reviewers to use the CONSORT-SPI

checklist to ensure that items are reported clearly.<sup>2,13</sup>

Free dissemination is key to widespread adoption. CONSORT-SPI guideline documents will be made available for online open-access in all journals publishing them. Several websites are also intended to host the guideline, including the CONSORT Group, EQUATOR Network, and CONSORT-SPI websites. An online discussion page or forum for feedback will allow for the guidelines' continual development. This webpage will be used to discuss new, relevant evidence related to social and psychological intervention trials, and to ask the wider scientific community to provide feedback on their experiences of using the guideline.<sup>20</sup>

Presentations have also been a key aspect of the initial dissemination and implementation strategy. The Project Executive, the IAG, and others at the consensus meeting have been and will continue to present the guideline at conferences, webinars, and meetings. Example conferences include those run by EQUATOR Network, the Guidelines International Network, and Cochrane and Campbell Collaborations. To increase uptake and buy-in, distribution through relevant organisations and practitioner networks is important. IAG members and consensus meeting participants will be asked to identify and present to the most influential professional bodies and organisations within their respective fields.

### ***2.3 Materials: CONSORT-SPI Consensus Meeting Discussion***

Two sessions on the final day of the consensus meeting related to knowledge translation activities and specifically sought feedback from journal editors in attendance and then from the group more generally. The DPhil candidate developed a presentation for each of these sessions, which were each led by a different discussant (P3 and P17, respectively). The presentations were based on the initial dissemination and

implementation plan from the project's grant application as well as strategies from the reporting guideline literature discussed above.<sup>3</sup> Meeting participants were informed that the goal of these two sessions was to optimise the plan for promoting use of the resultant CONSORT-SPI guidance, with the hope that members of the group would commit to specific efforts to this end.<sup>18</sup> Conversations revolved around the strategy for producing guideline documents, as well as ideas for disseminating and implementing these documents.<sup>3</sup> Common topics included insights from previous reporting guideline projects, ideas to better operationalise the project's existing dissemination and implementation strategies, and the generation of future knowledge translation activities.

To inform the dissemination and implementation plan in this chapter, the DPhil candidate conducted a directed content analysis of conversation during the two dissemination and implementation sessions.<sup>21</sup> The purpose of this analysis was to extend the theoretical frameworks detailed below to the area of reporting guidelines by identifying and categorising all potential suggestions for dissemination and implementation techniques according to these tools. As with the meeting sessions analysed in Chapter 5, the discussions at these two sessions were audio-recorded, transcribed verbatim by the DPhil candidate, and edited to remove personal identifiers.<sup>22</sup> As opposed to the approaches of previous thesis chapters,<sup>23</sup> coding of text in this chapter was directed by an existing coding scheme based on conceptual domains identified in the theoretical frameworks discussed below. The DPhil candidate first read the transcripts of these two sessions multiple times in order to become thoroughly situated with the data. The author then developed an initial set of coding categories based on key concepts from the theoretical frameworks, followed by line-by-line reading of the transcripts in which all text judged relevant to theoretical domains and behaviour change techniques was highlighted. Subsequently, all highlighted passages were assigned to a coding category and organised into lists of non-redundant

barriers and enablers of guideline use or knowledge translation strategies.

## **2.4 Procedure**

### *2.4.1 Step 1. Who Needs to Do What, Differently?*

The first step in developing the implementation strategy involved identifying who needs to do what differently to increase the chances of the CONSORT-SPI guideline being used and implemented. Namely, this involves specifying what are the primary behaviours, as well as any linked behaviours, that the implementation intervention is aiming to change. Those performing each of these behaviours, in addition to when and where they perform these behaviours, need to be identified.

### *2.4.2. Step 2. Using a Theoretical Framework, Which Barriers and Enablers Need to be Addressed?*

Following the identification of targeted behaviours and those performing them, theoretical frameworks need to be chosen that are most likely to inform the pathways of behaviour change. Several frameworks exist to assist in the development of implementation strategies for clinical guidelines and empirically supported interventions. Based on their prevalence in the implementation science literature, the following frameworks were used: the Guideline Implementability Framework (GIF),<sup>24,25</sup> the Promoting Action on Research in Health Services (PARiHS) Framework,<sup>26</sup> Consolidated Framework for Implementation Research (CFIR),<sup>27</sup> and Theoretical Domains Framework (TDF).<sup>12</sup> Each of these frameworks is based on surveys of relevant literature for barriers and facilitators to implementing clinical guidelines or interventions. Many of the frameworks overlap considerably in terms of their constructs and suggestions, though each has its own particular strengths and nuances. Each framework will be introduced in turn

below to briefly summarise the focus of each. The domains of each framework judged by the DPhil candidate to be relevant to barriers and enablers for the CONSORT-SPI dissemination and implementation strategy have been combined into one list in order to identify the malleable mechanisms to target.

*Guideline Implementability Framework.* The Guideline Implementability Framework (GIF)<sup>24,25</sup> aims to help guideline developers modify guidelines to support the various types of guideline use corresponding to different users. The framework was developed by reviewing implementation science literature and then applying insights to “implementability elements” used by a sample of publicly available, high-quality clinical guidelines. The final framework involves 22 elements that could improve guideline use, organised in the domains of guideline adaptability, usability, relevance, validity, applicability, communicability, resource implications, implementation, and evaluation. Examples include different versions of the guideline depending on the user and purpose of use, summaries of any evidence and recommendations in the guideline, potential resource implications of guideline use, and supporting information on how to promote and monitor guideline use.

*The PARiHS Framework.* The Promoting Action on Research in Health Services (PARiHS) framework is a widely cited conceptual framework for implementation strategies in health services.<sup>26</sup> Using this framework, implementers are prompted to consider three key interacting elements that impact successful implementation (SI) of practices: the evidence (E), the context (C), and facilitation of use (F). The general framework has been revised to better help specific research teams focused on the implementation of a targeted evidence-based practice, such as a reporting guideline.

*Consolidated Framework for Implementation Research.* The Consolidated Framework for Implementation Research (CFIR) was developed for interventions found effective in health services research that did not translate into meaningful outcomes in real-world practice.<sup>27</sup> CFIR provides an overarching typology for promoting effective implementation of interventions—such as clinical and reporting guidelines—and also developing knowledge about successful implementation strategies across populations and contexts. After synthesising constructs across published implementation theories, the CFIR consists of several constructs for implementation teams to consider within the five major domains: intervention characteristics, the inner setting of targeted organisations, the outer setting surrounding targeted organisations, the characteristics of individuals involved in implementation, and the process of implementation.

*Theoretical Domains Framework.* Another framework with significant theoretical and empirical support is the Theoretical Domains Framework (TDF): an integrative framework of theories of behaviour change, recently re-organised into 14 theoretical domains and 84 component constructs.<sup>9</sup> The aim of the Theoretical Domains Framework (TDF) is to simplify and comprehensively integrate behaviour change theories into a single framework to assess implementation problems and inform intervention design.<sup>12</sup> Over ten studies have already employed the TDF in exploratory qualitative methods to identify barriers and facilitators to guideline uptake for intervention design, and it has been used in combination with a framework for developing behavioural change interventions<sup>28</sup> in a variety of contexts to inform and address implementation problems.<sup>10</sup>

### *2.4.3 Step 3. Which Intervention Components Could Overcome the Modifiable Barriers and Enhance the Enablers?*

Once the modifiable barriers and facilitators have been identified, behaviour change techniques and their modes of delivery need to be identified. The proposed behaviour change techniques and modes of delivery are based on what was likely to be feasible and implemented as a coordinated implementation strategy.<sup>10</sup> The Behaviour Change Wheel<sup>28</sup> facilitated matching of barriers and enablers to knowledge transfer techniques identified from previous guidance and the initial CONSORT-SPI strategy. The Behaviour Change Wheel<sup>28</sup> aims to improve the design and implementation of evidence-based practices by improving the behavioural change interventions that make up dissemination and implementation activities. It does this by appropriately characterising interventions and then linking analysis of implementation to the targeted behaviour. At its core, the hub of the Behavioural Change Wheel is based on a “behaviour system” involving capability, opportunity, and motivation—the COM-B System. Around the hub of this system are positioned nine intervention functions that aimed at addressing deficits in these conditions, around which are placed seven categories of policy that could enable these interventions to occur. Facilitators and barriers from theoretical frameworks were mapped onto the COM-B System and matched with implementation intervention functions and policies using previously published categorisations.<sup>12</sup>

### *2.4.4 Step 4. How Can Behaviour Change Be Measured and Understood?*

While the face validity of the guideline in changing reporting behaviours is important, evaluating the actual impact of the guidelines on reporting quality is needed. The evaluation plan should involve a goal setting process that matches the actors, barriers and enablers, and techniques outlined above with specific, time-based targets that are

measurable, achievable, and realistic.<sup>12</sup> Both quantitative and qualitative feedback about the progress and quality of implementation can help the project team assess its impact as well as reflect on the utility of the guideline with future updates in mind.<sup>27</sup> Not only can such information indicate if the guideline is having its intended effect, it can also be used as audit and feedback to further increase uptake. Methods for assessing behaviour change and the mediators of change need to be selected that are reliable, valid, and feasible to evaluate.

### **3. Results**

#### ***3.1 Step 1. Who Needs to Do What, Differently?***

The primary target behaviour is the reporting of social and psychological intervention RCTs in journal articles, reports, and other document formats such as funding applications. As such, authors of trials need to use the CONSORT-SPI checklist and E&E to improve the accuracy, comprehensiveness, and transparency of information that they provide in trial reports. In addition to authors, peer-reviewers need to use the CONSORT-SPI checklist and E&E as tools for assessing the quality of information in reports of social and psychological intervention trials. These peer-reviewers could be serving journals or other organisations that conduct peer-reviews of trial reports like funding organisations, or they may be journal editors or funding agency staff themselves.

Several secondary behaviours should also be targeted that relate to those organisations envisioned to implement CONSORT-SPI. Journals publishing social and psychological intervention trials are intended to endorse and enforce adherence to the CONSORT-SPI documents by trial authors and peer-reviewers. Funding agencies are hoped to do the same with applications for funding social and psychological intervention trials as well as final reports on trials they agree to fund. Faculty and others who teach

graduate students or early career researchers need to incorporate reporting guidelines and reporting standards into their lessons.

Lastly, the scientific community at large needs to be more proactive in its support of reporting guidelines. Any stakeholders involved in some way in the production of research, developing practice guidelines or policies, and consuming social and psychological services should place greater pressure on those conducting and publishing social and psychological intervention research to follow reporting standards. These broader members of the scientific community may consist of editorial groups, publishing companies, research or professional societies, university departments or research centres, government bodies, or those involved in grading evidence for policy, such as the US Substance Abuse and Mental Health Services Administration's National Registry of Evidence-based Programs and Practices.

The disciplines, professions, and geographic locations of those being targeted should also be considered, particularly with the timing of particular components of the implementation strategy in mind. For example, some fields may conduct and report RCTs more frequently than others.<sup>16</sup> Several participants at the CONSORT-SPI consensus meeting also noted that research communities in the United States and United Kingdom are ahead of continental Europe and low- and middle-income countries in terms of using reporting standards and guidelines, and that those working for government who conduct intervention trials may have different drivers for and attitudes toward adherence to reporting guidelines than researchers working in academic settings. The CONSORT-SPI dissemination and implementation strategy therefore needs to reflect that some areas may be more ready than others to change behaviours due to the maturity of conducting trials in the area.

### ***3.2 Step 2. Using a Theoretical Framework, Which Barriers and Enablers Need to be Addressed?***

Potential barriers and enablers to implementing CONSORT-SPI relate to the nature of the guideline itself as well as individual-, organisational-, and system-level factors.

#### *3.2.1 Barriers and Enablers of the Guideline Itself*

These barriers and enablers have to do with how the guideline itself is presented and organised (see Table 1). Firstly, accommodation to use or implement the guideline may relate to how clearly the guideline documents themselves indicate the purpose and rationale of the guideline as well as the costs, resources, and competencies required.<sup>24</sup> In particular for CONSORT-SPI, this accommodation may depend on how explicit the guideline notes that the objective of the guideline is to improve reporting quality, and who the direct users and indirect implementers are. Secondly, the soundness of the evidence and recommendations within the guideline may also prevent or facilitate guideline use. The perceived validity could relate to the number of references provided for evidence and the total number of recommendations within the guideline.<sup>24</sup> In addition, several aspects of the guideline could make it difficult to apply. The format of both the recommendations and the evidence underpinning them,<sup>24</sup> as well as the complexity of the guideline and the quality of how it is designed and packaged,<sup>27</sup> could either help or hinder the ease of guideline use. The nature of contextual information provided to help the reader interpret and apply recommendations to individual trials impacts guideline use as well.<sup>24</sup> Lastly, the CONSORT-SPI team should consider the suitability of guideline documents for different users or purposes (e.g., trial authors, peer-reviewers), as well as the ability to apply the guideline at point of use (e.g., submitting a trial report).<sup>24</sup>

**Table 1. Potential Barriers and Enablers of the Guideline Itself**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Clarity of guideline purpose, costs and resources needed, and competencies to use	GIF	Accommodation: Objective	Opportunity: Physical
		Accommodation: Users	Opportunity: Physical
Perceived soundness of evidence and recommendations	GIF	Validity: Number of references	Opportunity: Social
		Validity: Number of recommendations	Opportunity: Social
Difficulty of using the guideline	GIF	Usability: Recommendation format	Opportunity: Physical
		Usability: Evidence format	Opportunity: Physical
	CFIR	Intervention Characteristics: Complexity	Opportunity: Physical
		Intervention Characteristics: Design quality and packaging	Opportunity: Physical
Contextual information to help reader interpret and apply recommendations to individual trials	GIF	Applicability: Individualisation	Opportunity: Physical
Versions of guideline tailored for different users or purposes	GIF	Adaptability: Alternate versions	Opportunity: Physical
Ability to apply guidelines at point-of-use	GIF	Implementation: Tools	Opportunity: Physical

### 3.2.2 Individual-Level Barriers and Enablers

These barriers and enablers have to do with individuals who will actually use the guidelines, such as authors, journal editors, and peer-reviewers. They typically represent psychological limitations or strengths of individuals to use the guideline.<sup>12</sup>

*Factors before using the guideline.* Several barriers and enablers may occur before an individual decides to use CONSORT-SPI (see Table 2). Individuals' knowledge about the scientific rationale for CONSORT-SPI, procedural knowledge on how to use the guidance, and the understanding of costs of and resources needed for guideline use should be considered.<sup>12,24</sup> Individuals may also associate specific outcomes with guideline use, which can serve as barriers or enablers depending on the probability that individuals think they will occur, how quickly they may occur, and how much individuals value them or find them important.<sup>12</sup> The CONSORT-SPI implementation strategy also needs to address individuals' attitudes toward the guideline, particularly whether they see the guideline as being developed and imposed on them by someone outside of their own social or professional group.<sup>27</sup> All of the above may indicate individuals' perceptions that the guideline is useful,<sup>26</sup> their optimism that use of the guideline is for the best,<sup>12</sup> and ultimately their readiness to use the guideline.<sup>27</sup>

*Within-person factors while using the guideline.* After individuals decide to try and use CONSORT-SPI, other barriers and enablers may arise due to characteristics of the persons attempting to use the guideline (see Table 3). Successful attempts will be partly determined by individuals' perceptions of ease of using the guideline, their beliefs in their general skill set and abilities as they relate to using reporting guidelines, their beliefs in their capability to specifically use CONSORT-SPI for a given purpose, and their actual abilities and competence to use the guideline.<sup>12,27</sup> These are of particular concern when the demands placed on an individual by the guideline are greater than that individual's abilities

**Table 2. Potential Individual-Level Barriers and Enablers before Using CONSORT-SPI**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Individual degree of knowledge about the rationale underpinning the guideline	TDF	Knowledge: Scientific rationale	Capability: Psychological
Individual degree of knowledge about how to use the guideline	TDF	Knowledge: Procedural knowledge	Capability: Psychological
Clarity to individual users of costs of guideline use and resources needed	GIF	Accommodation: Costs	Capability: Psychological
Specific outcomes associated by the individual with guideline use	TDF	Beliefs about consequences: Outcome expectancies	Motivation: Reflective
Characteristics of the specific outcomes that individuals associate with guideline use	TDF	Beliefs about consequences: Characteristics of outcome expectancies	Motivation: Reflective
Potential perception by individual stakeholders that guideline is "foreign"	CFIR	Intervention characteristics: Intervention source	Opportunity: Social
Individuals' attitudes toward the guideline	CFIR	Characteristics of individuals: Knowledge and beliefs about the intervention	Motivation: Automatic
Individual perceptions of guideline utility	PARiHS	Evidence and EBP characteristics	Motivation: Reflective
Individual confidence that use of guideline would be for the best	TDF	Optimism	Motivation: Reflective
Characterisation of an individual's phase in their progression toward skilled, enthusiastic, and sustained use of the guideline	CFIR	Characteristics of individuals: Individual stage of change	Capability: Psychological

**Table 3. Potential Within-Individual Barriers and Enablers while Using CONSORT-SPI**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Individual's perception of the ease or difficulty of using the guideline	TDF	Beliefs about capabilities: Perceived behavioural control	Motivation: Reflective
Individual's belief in existing repertoire of skills and ability as they relate to using the guideline	TDF	Beliefs about capabilities: Professional confidence	Motivation: Reflective
Individual's perceived capacity to use guideline effectively	CFIR; TDF	Self-efficacy	Motivation: Reflective
Individual ability and competence to use the guideline	TDF	Skills: Ability and Competence	Capability: Psychological
Demands placed on an individual by guideline use are greater than that individual's mental abilities	TDF	Memory, attention and decision processes: Cognitive overload / tiredness	Capability: Psychological
Individual's ability to retain information about use of guideline from previous experience using it	TDF	Memory, Attention and Decision Processes: Memory	Capability: Psychological
Internal feelings or states that occur when guideline is used	TDF	Emotion: Positive / negative affect, stress, anxiety, burn-out	Motivation: Automatic
Ability of individuals to practice skills needed to use the guideline	TDF	Skills: Practice	Capability: Physical
Potential need to discontinue behaviours automatically activated during manuscript writing	TDF	Behavioural regulation: Breaking habit	Capability: Psychological
Whether an individual keeps record of guideline use in connection with improved reporting	TDF	Behavioural regulation: Self-monitoring	Capability: Psychological

to adhere to them.<sup>13</sup> Several individual responses during guideline use may also need to be addressed, such as stress, apprehension due to anticipation of impending negative consequences, or exhaustion that leads to lowered performance of guideline use.<sup>13</sup> The implementation plan should also look to attend to whether individuals have the ability or opportunities to practice the skills they need to use CONSORT-SPI, and they keep records of improved reporting in connection with their guideline use.<sup>13</sup> Lastly, there is a potential need to discontinue behaviours or break habits automatically activated during manuscript writing and submission that would forgo individuals' use of guidelines.<sup>13</sup>

*Individual guideline use in context.* While characteristics of the individual are important to consider, the CONSORT-SPI implementation plan should also take into account individuals' relationship to the surrounding context (see Table 4). Several potential barriers and enablers relate to individuals' relationships with organisations seeking to implement CONSORT-SPI within the organisation context. The implementation strategy should also consider how individuals perceive their relationship with and commitment to any organisation seeking to implement CONSORT-SPI, such as a given journal, editorial group, research society or funding organization.<sup>28</sup> This may include an estrangement from or dissatisfaction with these organisations or professional groups by the individuals.<sup>13</sup> Should a positive relationship exist, actual use of CONSORT-SPI by individuals may depend on how they perceive use of the guideline as coherent with their identity within the organization, and their confidence to carry out professional responsibilities related to guideline use (e.g., complete peer-review in a timely manner).<sup>13</sup>

In addition to their relationship to an organisation, individuals may or may not use CONSORT-SPI depending on the advantages of doing so in comparison to other individuals in their social or professional group,<sup>27,28</sup> or how individuals believe their abilities and performance in guideline adherence compares to others.<sup>13</sup> Their stability of

**Table 4. Potential Contextual Barriers and Enablers during Individuals' Use of CONSORT-SPI**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
How individuals perceive their relationship with and commitment to a given organisation	CFIR	Characteristics of individuals: Individual identification with organisation	Motivation: Reflective
Estrangement from or dissatisfaction with professional group implementing guideline	TDF	Social influences: Alienation	Opportunity: Social
Confidence of individuals to carry out professional responsibilities related to guideline use	TDF	Social/professional role and identity: Professional confidence	Motivation: Reflective
Coherence of guideline use with individuals' professional identity in a social or work setting	TDF	Social/professional role and identity: Professional identity	Motivation: Reflective
Relative advantage to individuals of using guideline	PARiHS	Evidence and EBP Characteristics	Motivation: Reflective
	CFIR	Intervention Characteristics: Relative advantage	Motivation: Reflective
How individuals might evaluate their attitudes, abilities, or performance in guideline adherence relative to others in a community	TDF	Social influences: Social comparisons	Opportunity: Social
Individual's resolve to continue guideline use in spite of disturbing influences	TDF	Intentions: Stability of intentions	Motivation: Reflective
Accessibility of information to trial authors and peer-reviewers that supports guideline use	GIF	Communicability	Opportunity: Physical
Degree to which an individual has exemplars or models to follow or copy	TDF	Social influences: Modelling	Opportunity: Social

intentions to use the guideline despite external challenges is key to consider,<sup>12</sup> which could be influenced by accessibility of information they have to support use of the guideline,<sup>25</sup> and the degree to which individuals have exemplars to model.<sup>12</sup>

### 3.2.3 Organisational-Level Barriers and Enablers

In addition to barriers and enablers of individual guideline use, organisations may also face hindrances in implementing CONSORT-SPI in their organisational context that need to be addressed by the CONSORT-SPI implementation strategy.

*Factors before implementing the guideline.* As with individual guideline use, several barriers and enablers may occur before an organisation decides to implement CONSORT-SPI (see Table 5). Firstly, the costs to and resources needed by the organisation to implement the guideline are important to consider.<sup>24</sup> Organisations will likely make decisions on implementation depending on the specific outcomes they associate with guideline use and the probability important outcomes will be attained in a timely manner.<sup>12,26</sup> Such considerations should include organisational sense of the potential negative consequences of implementing guidelines and the anticipated regret of guideline implementation in light of these consequences.<sup>12</sup> The degree to which these outcomes relate to the ultimate goals of an organisation (e.g., funding informative trials) may dictate guideline implementation.<sup>12</sup> Organisations will be less likely to use CONSORT-SPI if they see it as an imposition from an entity foreign to their professional or social group,<sup>27</sup> or if use of reporting guidelines by individuals within the organisation does not coincide with organisational membership identification or the culture of an organisation.<sup>12</sup> The above factors could impact an organisation's view that implementing CONSORT-SPI is for the best, their motivations for implementing it, and their confidence in being able to do so effectively.<sup>12</sup>

**Table 5. Potential Organisational-Level Barriers and Enablers before Implementing CONSORT-SPI**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Clarity to organisations of costs of guideline implementation and resources needed	GIF	Accommodation: Costs	Opportunity: Physical
Specific outcomes associated by the organisation with guideline use	TDF	Beliefs about consequences: Outcome expectancies	Motivation: Reflective
Characteristics of the specific outcomes that organisations associate with guideline use	TDF	Beliefs about consequences: Characteristics of outcome expectancies	Motivation: Reflective
Organisational perceptions of guideline utility	PARiHS	Evidence and EBP Characteristics	Motivation: Reflective
Organisational sense of the potential negative consequences of implementing guideline	TDF	Beliefs about consequences: Anticipated regret	Motivation: Reflective
Proximity of organisational goals after implementation, in relation to goals of guideline	TDF	Goals	Motivation: Reflective
Potential perception by organisations that guideline is "foreign"	CFIR	Intervention characteristics: Intervention source	Opportunity: Social
Degree that guideline use matches characteristics by which groups or organisations identify membership	TDF	Social influences: Group identity	Opportunity: Social
Compatibility of guideline use with values, attitudes, beliefs, and customs of an organisation	TDF	Environmental context and resources: Organisational culture /climate	Opportunity: Physical
Organisational confidence that guideline use is for the best	TDF	Optimism	Motivation: Reflective
Stage of change for organisation's implementation of guideline	TDF	Intentions: Transtheoretical model and stages of change	Motivation: Reflective
Confidence of organisation to effectively enforce guideline adherence	TDF	Social/professional role and identity: Professional confidence	Motivation: Reflective

*Within-organisation factors while using the guideline.* Once an organisation has decided to implement CONSORT-SPI, several factors could facilitate or hinder the success of its implementation (See Table 6). Proper implementation will likely depend on the compatibility of the guideline with structural characteristics and culture of an organisation, such as its age, size, maturity in the area of trials, and values towards guideline use.<sup>27</sup> Amongst an organisation's members, the group norms about guideline implementation and the perceived coherence of guideline use with their roles in the organisation as well as the identity of the organisation need to be addressed.<sup>12</sup>

The degree of leadership support for and organisational commitment to implementation of CONSORT-SPI is likely to influence these potential barriers related to group norms and professional roles.<sup>12,27</sup> The amount of information, physical resources, and human resources made available by the organisation to individuals for using CONSORT-SPI also needs to be considered.<sup>12,27</sup> Factors of the organisation that may cause stress when using the guideline should be addressed,<sup>12</sup> such as too strict of timeframes for completing CONSORT-SPI checklists and incorporating them in peer-reviews, particularly for new users. Social influences to guideline use within an organisation might be relevant to these factors, such as discouragement from peers to use the guideline.<sup>12</sup> Whether organisations have the capacity to promote guideline use despite resistance probably impacts whether these influences can serve as barriers.<sup>12</sup> A facilitator of this is the capacity to develop a plan created in advance by the organisation of when, where, and how to implement the guideline, such as the ability to pilot guideline use on a small scale, guarantee particular outcomes of guideline use by individuals within the organisational context, and keep record of implementation and reporting quality.<sup>12,27</sup>

*Organisational guideline implementation in context.* While characteristics of the organisation itself need to be addressed, the role of the organisation in relation to the wider

**Table 6. Potential Within-Organisation Barriers and Enablers while Using CONSORT-SPI**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
The structure, age, maturity, size, and social networks within an organisation	CFIR	Inner setting: Structural characteristics, Networks & communications	Capability: Physical
The perceived compatibility of the guideline with an organisation's norms and values	CFIR	Inner setting: Culture, Implementation climate	Opportunity: Social
Degree to which guideline implementation is held to be correct or acceptable by a given group	TDF	Social influences: Group norms	Opportunity: Social
Coherence of guideline use with member roles and organisational identity	TDF	Social/professional role and identity: Professional role, Group identity	Motivation: Reflective
The nature and quality of social networks and communications within an organisation	CFIR	Inner setting: Networks & communications	Opportunity: Social
Degree of leadership support for and organisational commitment to implementation	CFIR, TDF	Social/professional role and identity: Leadership, Organisational commitment	Motivation: Reflective
Information, commodities and human resources made available by the organisation to individuals	CFIR; TDF	Environmental context and resources: Resources / material resources	Opportunity: Physical
Factors of the organisation that cause stress when individuals use the guideline	TDF	Environmental context and resources: Environmental stressors	Opportunity: Physical
The capacity of an organisation to promote guideline use, even when individuals try to resist	TDF	Social influences: Power	Opportunity: Social
Existence and quality of plan created in advance of when, where and how to implement	CFIR; TDF	Goals: Implementation intention	Motivation: Reflective

scientific community needs to also be respected (see Table 7). The support of leaders in relevant research societies could facilitate implementation of the guideline,<sup>26</sup> though estrangement from or dissatisfaction with societies, other groups, and wider aspects of the scientific community in general may act as a barrier.<sup>12</sup> Organisations may also be influenced to implement the guideline well if a relative advantage would ensue or if there is competitive pressure from other organisations to implement it.<sup>26,27</sup> An organisation's view of their ability to or performance in enforcing the guideline more successfully than other organisations could modify the influence of these competitive factors.<sup>12</sup>

Organisations may be hesitant to implement the guideline, though, if they take a view that guideline enforcement may overstep professional boundaries.<sup>12</sup> Coordinating with other groups, or incentives from external policy or regulatory bodies, may lessen such worries about overstepping boundaries.<sup>27</sup> The existence of exemplars or models to follow might also enable organisations to keep to intentions to implement the guideline effectively.<sup>12</sup>

### *3.2.4 System-Level Barriers and Enablers*

In addition to considering drivers of guideline use by individuals and guideline implementation by organisations, the wider support of reporting guidelines and reporting standards in the scientific community may depend on several potential barriers and enablers as well.

*Factors before support of reporting guidelines.* Many scientific communities within the social and behavioural sciences may not yet support the idea of reporting guidelines and reporting standards, with several possible barriers and enablers for getting them to support these (see Table 8). Firstly, certain scientific communities might not even be aware of reporting standards or guidelines, or at least familiar with the rationale underpinning

**Table 7. Potential Contextual Barriers and Enablers during Organisational Implementation of CONSORT-SPI**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Support of research society leaders	PARiHS	Contextual readiness for targeted EBP implementation	Opportunity: Social
Estrangement from or dissatisfaction with wider scientific community by groups or organisations	TDF	Social influences: Alienation	Opportunity: Social
Relative advantage to organisations of implementing guideline	PARiHS	Evidence and EBP Characteristics	Motivation: Reflective
	CFIR	Intervention characteristics: Relative advantage	Motivation: Reflective
Competitive pressure on organisation to implement guideline	CFIR	Outer setting: Peer pressure	Opportunity: Social
How organisations might evaluate their ability or performance in guideline enforcement relative to others in a community	TDF	Social influences: Social comparisons	Opportunity: Social
Degree to which implementation and enforcement oversteps professional boundaries	TDF	Social/professional role and identity: Professional boundaries	Motivation: Reflective
Degree to which organisation coordinates implementation with external organisations	CFIR	Outer setting: Cosmopolitanism	Opportunity: Physical
External incentives for an organisation to use guidelines	CFIR	Outer setting: External policy & incentives	Opportunity: Physical
Organisation's resolve to continue guideline implementation in spite of disturbing influences	TDF	Intentions: Stability of intentions	Motivation: Reflective
Degree to which an organisation has exemplars or models to follow or copy	TDF	Social influences: Modelling	Opportunity: Social

**Table 8. Potential Community-Level Barriers and Enablers before Supporting Reporting Guidelines**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Scientific community degree of knowledge about rationale for reporting standards/guidelines	TDF	Knowledge: Scientific rationale	Capability: Psychological
Clarity to behavioural and social sciences of costs and resources needed to support guidelines	GIF	Accommodation: Regulatory	Capability: Psychological
Specific outcomes associated by the scientific community with support of standards/guidelines	TDF	Beliefs about consequences: Outcome expectancies	Motivation: Reflective
Characteristics of the specific outcomes that community associates with guideline support	TDF	Beliefs about consequences: Characteristics of outcome expectancies	Motivation: Reflective
Relative advantages of supporting reporting standards and guidelines	PARiHS	Evidence and EBP characteristics	Motivation: Reflective
	CFIR	Intervention characteristics: Relative advantage	
Scientific community perceptions of reporting standards' and guidelines' utility	PARiHS	Evidence and EBP characteristics	Motivation: Reflective
Characteristics of the end state where reporting standards and guidelines might be normalised	TDF	Goals	Motivation: Reflective
Potential perception by scientific community that guidelines and standards are "foreign"	CFIR	Intervention characteristics: Intervention source	Opportunity: Social
Scientific community's confidence that support of reporting standards and guidelines is for the best	TDF	Optimism	Motivation: Reflective
Stage of change for community's support of guidelines	TDF	Intentions: Transtheoretical model and stages of change	Motivation: Reflective

them.<sup>13</sup> As such, the costs and resources needed to support reporting guidelines and the standards therein—such as trial registries and data repositories—need to be clear.<sup>25</sup>

In addition to knowledge about reporting guidelines, the specific outcomes associated with reporting standards and guidelines by scientific communities are important to consider, given that these may prevent or facilitate support of guidelines given the characteristics of associated outcomes.<sup>13</sup> For example, certain scientific communities might find relative advantage<sup>27</sup> in supporting reporting guidelines and standards if doing so recruits higher- quality research studies to journals and academic departments within their area as opposed to similar areas in the behavioural and social sciences. These relative advantages may depend in part on a community's perceptions of the overall utility of reporting guidelines and standards,<sup>27</sup> as well as the end game or goal to which reporting standards might be aimed.<sup>13</sup> One potential barrier is the belief by scientific communities that reporting standards and guidelines might be seen as imposed by biomedical sciences, as most to date have been developed for health research more specifically than behavioural and social science research.<sup>1,17,28</sup> All the above may be part of a community's general belief that supporting the implementation and use of reporting guidelines within their area is for the best, indicating their readiness to change behaviours regarding research reporting practices.<sup>13</sup>

*Within-community factors while supporting reporting guidelines.* Once a scientific community has demonstrated general support of reporting guidelines, several factors could serve as barriers or enablers to successful support of individuals and organisations within the community to implementing and use reporting guidelines (See Table 9). A community's identity and social norms as they relate to supporting reporting guidelines need to be taken into account.<sup>13</sup> In particular, support for guideline implementation and use may depend on the degree to which a scientific community can consciously and

**Table 9. Potential Within-Community Barriers and Enablers while Supporting Reporting Guidelines**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Coherence of supporting reporting standards and guidelines with scientific community identity	TDF	Social/professional role and identity: Group identity	Motivation: Reflective
Socially determined consensual standards that indicate the typicality and approval of the use and implementation of reporting standards and guidelines within a scientific community	TDF	Social influences: Social norms	Opportunity: Social
Degree of conscious maintenance versus disagreement or confrontation amongst groups within a community regarding reporting guidelines	TDF	Social influences: Intergroup conflict, Group conformity	Opportunity: Social
Capacity of a scientific community to influence its members to implement and use guidelines, even when individuals or groups try to resist	TDF	Social influences: Power	Opportunity: Social
Degree of commitment by scientific community and its leaders to reporting standards	TDF; PARiHS	Social/professional role and identity: Organisational commitment, Leadership	Motivation: Reflective
Whether a scientific community keeps record of use of reporting standards and guidelines in connection with improved reporting	TDF	Behavioural regulation: Self-monitoring	Capability: Psychological

harmoniously come together to raise reporting standards, or whether there is likely to be fragmentation from disagreement amongst groups within a community regarding reporting guidelines.<sup>12</sup> The power or capacity of a scientific community to influence its members to implement and use guidelines despite such resistance needs to be addressed for reporting guideline implementation strategies.<sup>12</sup> One potential facilitator is the degree to which the leaders of scientific communities support commitment of the community to implement reporting standards.<sup>12,26</sup> Another facilitator would be the ability of a scientific community to keep record of the reporting quality of its research, particularly in conjunction with movements to incorporate reporting standards and guidelines into community publishing practices.<sup>12</sup>

*Wider factors influencing reporting guideline support.* Several contextual factors may influence whether scientific communities can effectively support reporting guideline use and implementation (see Table 10). Firstly, other scientific communities or systems related to research—such as government allocation of funds for research—may exert influence on a scientific community's efforts to support reporting standards and guidelines.<sup>12</sup> For example, a distinct, prominent event may occur outside of a specific scientific community that directly impacts the support for reporting standards and reporting guidelines in a community, such as the new US National Institutes of Health plans to enhance reproducibility of research.<sup>29</sup> Other factors external to the scientific community may cause stress for organisations implementing reporting guidelines and individuals trying to using them,<sup>12</sup> such as potential resistance from publishing companies to support changes in journal editorial systems to facilitate guideline use.<sup>26</sup> The impact of any such influences that are social rather than regulatory in nature may depend on the degree to which a scientific community is estranged from or dissatisfied with the other communities or systems exerting said influence.<sup>12</sup>

**Table 10. Potential Contextual Barriers and Enablers during Community Support of Reporting Guidelines**

<b>Barrier or Enabler</b>	<b>Framework</b>	<b>Theoretical Domain</b>	<b>COM-B Component</b>
Exertion of influence of another system on the scientific community to support reporting standards and guidelines	TDF	Social influences: Social pressure	Opportunity: Social
Distinctive, prominent or otherwise significant occurrences related to use and implementation of reporting standards and guidelines	TDF	Environmental context and resources: Salient events / critical incidents	Opportunity: Physical
Estrangement from or dissatisfaction with other communities/systems by a scientific community	TDF	Social influences: Alienation	Opportunity: Social
Factors external to the scientific community that cause stress for organisations implementing the guideline and individuals using it	TDF	Environmental context and resources: Environmental stressors	Opportunity: Physical
Supportiveness of publishing culture for reporting standards and guidelines	PARiHS	Contextual readiness for targeted EBP implementation	Opportunity: Social

### ***3.3 Step 3. Which Intervention Components Could Overcome the Modifiable Barriers and Enhance the Enablers?***

A multi-faceted, multi-level strategy is needed to address the theoretically-informed barriers and enablers identified above.

#### *3.3.1 Guideline-Level Implementation Intervention Components*

These components of the CONSORT-SPI dissemination and implementation strategy target the organisation and presentation of CONSORT-SPI guideline documents. The CONSORT-SPI Extension Statement, Checklist, and E&E need to clearly state in the introduction the purpose of CONSORT-SPI as well as its targeted users and beneficiaries so that interested readers can understand the rationale underpinning the guideline and for whom the guideline is intended. To demonstrate the perceived validity of the guideline recommendations and evidence underpinning them, the checklist provides distinct, clear, and tabulated recommendations, with referenced evidence discussed narratively throughout the E&E. The prestige of the authorship group should also signal the face validity of and experience going into the guideline recommendations.

Furthermore, this checklist is minimal to reduce its burden and complexity for users, as long checklists can deter authors from using them and journal editors from enforcing them, particularly amongst those who see reporting items as part of larger bureaucratic “box-ticking” exercises. Checklist items will be crisply and unambiguously worded and organised intelligibly by sections of a manuscript to demonstrate their usefulness in writing trial manuscripts more easily and efficiently. A glossary of key terms will also be developed to ensure that their meanings are clear.

To ensure recommendation applicability and adaptability to users with various scientific backgrounds, recommendations in the E&E will be contextualised depending on

the different types of interventions and trials in the area, with various versions of the E&E adapted for different disciplines as well. Alternate versions of the checklist and flow diagram may also be produced for authors and peer-reviewers rather than for general consumption, such as an editable template that can be used during manuscript submission and review as compared to an unalterable PDF document.

Rather than simply upload them online on a project website, guidance documents will be submitted for publication in professional, peer-reviewed manuscripts, which should improve the quality of the “packaging” of the guidance. The Extension Statement and E&E documents will be published simultaneously with the Extension Statement to ensure the supporting evidence is available along with the checklist. A copy of the checklist will also be published in the E&E, and standalone versions of the checklist will direct authors to use the E&E, to address concerns that the E&E may not be used at all when using the checklist. These are intended to be published across a variety of high-impact, peer-reviewed journals in disciplines that may have separate “literatures” in order to increase the reach of the guideline, with one corresponding author to elicit journals’ interest, coordinate peer-reviews, and assist with copy-editing.<sup>3</sup> Members of the Project Executive, International Advisory Group, and consensus meeting may help to identify and approach the most appropriate journals in key disciplines to publish the guideline and to provide an editorial supporting the guideline.

Free dissemination is key to widespread use and implementation. Guideline documents will also be made open access so that they are free for all to access and reproduce with proper citation, as well as to allow for their publication on websites such as those run by the CONSORT-SPI Project Executive, the CONSORT Group, and the EQUATOR Network in order to increase their visibility and accessibility. Resources and demand pending, the guideline documents could also be translated into those languages

most likely to benefit, with the CONSORT-SPI Project Executive officially involved in these translations to ensure that they are done well.<sup>3</sup>

### *3.3.2 Individual-Level Implementation Intervention Components*

The following implementation intervention components target those individuals who will actually use the guideline documents to improve the reporting quality of social and psychological intervention RCTs: namely, the trial authors and peer-reviewers. The wider group involved in developing CONSORT-SPI can be involved in writing editorials, distributing messages through professional networks and societies' communication channels, and presenting at major conferences or meetings to raise awareness about CONSORT-SPI and its utility. Information on the potential efficiency of writing and peer-reviewing by using a rubric, as well as the higher quality of reports, may persuade authors and reviewers to use the guideline. Training programmes, workshops at conferences, videos publicly accessible online, or webinars providing instructions on how to use the website may help authors and peer-reviewers develop the skills needed to adhere to the checklist effectively. Faculty of graduate and post-doctoral programmes in particular may incorporate CONSORT-SPI into their curriculum so that researchers-in-training learn these skills early. In coordination with groups like the CONSORT Executive and the EQUATOR Network, the CONSORT-SPI team may also develop an online training course that leads to accreditation as an author or peer-reviewer certified to use reporting guidelines like CONSORT-SPI.

Comparison of behaviour techniques may prove useful, such as information about other individuals' approval of the guideline documents.<sup>30</sup> For example, presentation of the results of the Delphi process, the consensus meeting votes, and the respected authors of the guidelines in the documents themselves as well as on the project website may lead

individuals to be more in favour of using CONSORT-SPI. In addition, exemplar trials fully adherent to CONSORT-SPI may be published in the same journal issue as the CONSORT-SPI guideline documents to provide models for others to follow. The CONSORT-SPI website can also keep a list of exemplar trials for authors and peer-reviewers to refer.

### *3.3.3 Organisational-Level Implementation Intervention Components*

The following components of the dissemination strategy target those organisations that are in a position to execute implementation policies within their organisations requiring adherence to CONSORT-SPI. Examples of such organisations include specific journals, editorial groups, funding organisations, and research institutions. Firstly, the “public relations” techniques listed above—such as editorials, conference presentations, and organisational newsletters—could also highlight to organisations the expected gains to their organisations of enforcing use of CONSORT-SPI in relevant organisational activities. For example, manuscript submissions to journals and their subsequent publications may be of higher quality, possibly even with more citations and impact; funding organisations may similarly have better quality applications and greater return-on-investment when publications of their funded trials provide the information needed by users like practitioners and policy-makers. The project website in particular could signal to organisations all the journals, research organisations, and professional societies that have been involved in developing or have already indicated endorsement of the guideline.

One key strategy is recruiting opinion leaders within these organisations to serve as either formally appointed internal implementation leaders or as “champions” that help to ensure the commitment and accountability of an organisation with the implementation of CONSORT-SPI. These individuals should have formal or informal influence on the colleagues’ attitudes, beliefs, and behaviours within the organisational context, and would

be willing to take responsibility for and dedicate themselves to driving through the implementation of CONSORT-SPI within their organisation.<sup>27</sup> For example, they could be dedicated experts on a journal editorial board that assists other editors in maintaining a CONSORT-SPI implementation policy.<sup>12</sup> The CONSORT-SPI Project Executive could also help organisations recruit “external change agents” who are affiliated with an outside entity that formally influences or facilitates implementation of CONSORT-SPI within the organisation.<sup>27</sup> For example, journals that are run by research societies could elect a member of their editorial board to liaise with the research society in order to encourage the society’s support of implementing CONSORT-SPI within its journals.

Particularly if such programme champions exist, the CONSORT-SPI team can help these organisations institute their own implementation interventions within their settings. For example, organisations can collect information about reporting quality in their journal and adherence to guidelines by authors and peer-reviewers, and feed this information back to editorial staff periodically at meetings. In addition, members of the wider CONSORT-SPI team could help to identify and contact representatives from all organisations to endorse the guideline or incorporate CONSORT-SPI into more general reporting guidelines that they currently enforce (e.g., the American Psychological Association Journal Article Reporting Standards<sup>31</sup>). The meaning of endorsement should be explicit, such as clear directions in journals’ “Instructions to Authors” sections of their websites or peer-reviewer guidelines that state a CONSORT-SPI checklist should be completed in consultation with the E&E for all submissions of social and psychological intervention trials. Ideally, journal policy would have submissions without a checklist be returned to authors,<sup>2,32</sup> or use checklists to make decisions about publication, such as having submissions that are CONSORT compliant receive a greater prominence on the journal website or volume within which it is printed. If the “reporting guideline certification”

suggestion in the section above is enacted, organisations could require such certification by trial authors or peer-reviewers handling trial submissions. Representatives of social and behavioural science programmes within universities could be approached by the CONSORT-SPI team at conferences and asked if they would ensure that the guideline is included in curriculum. Research organisations could contractually require adherence to CONSORT-SPI in articles and final reports of social and psychological intervention trials that they fund.

### *3.3.4 System-Level Implementation Intervention Components*

The CONSORT-SPI team could be involved in other activities that seek to get communities within the social and behavioural sciences more generally supportive of reporting guidelines and standards like CONSORT-SPI. A key activity is continually engaging, attracting, and involving influential stakeholders in the social marketing and education on reporting guidelines in general.<sup>27</sup> For example, Associate Deans of Research within a discipline, who attend annual meetings, could develop strategies to disseminate information about reporting standards to their faculty and staff. Incorporating CONSORT-SPI and other reporting guidelines into the curriculum in social and behavioural science graduate programmes could lead to a generational recognition of the importance of reporting guidelines. The use of reporting guidelines could consequently become normative as this generation of graduate students progress in their scientific careers.

Editorial groups or larger research societies could endorse the guideline, making any possible competitive disadvantages of being early implementers of CONSORT-SPI go away, as well as establishing a new baseline standard of reporting for an entire area. Those organisations involved in grading the evidentiary status of social and psychological interventions—organisations like the What Works Clearinghouse, National Registry of

Evidence-based Programs and Practices, and Blueprints for Healthy Youth Development—could use reporting standards to underpin their criteria, or they could give interventions “extra credit” for their certifications if they report CONSORT-compliant trials.

The EQUATOR Network is an essential resource to getting scientific communities more broadly in favour of reporting guidelines and standards. For example, like the Cochrane Collaboration, the EQUATOR Network could continue to create a satellite network of regional hubs that spread awareness of reporting guidelines to scientific communities, given the work it has already done to clarify how to navigate the library of reporting guidelines in health and related research. They could continue to highlight events on the news section of their website or in their newsletters that might facilitate use of guidelines, such as information about movements like AllTrials with which reporting guideline developers could combine efforts. A grading or appraisal system for reporting guidelines themselves could also be developed to provide users of reporting guidelines assurance of their quality. Lastly, journals could be encouraged to move to online publishing, which challenges current publishing norms about word limits and facilitates more information to be included or linked in manuscripts.

#### ***4.4 Step 4. How Can Behaviour Change be Measured and Understood?***

Success of the CONSORT-SPI guideline primarily would be measured by improved quality of reporting social and psychological intervention trials. A rubric for reporting quality based on the finalised CONSORT-SPI checklist could be used to see how well manuscripts accurately, completely, and transparently provide information needed to critically appraise, apply, and replicate trials. The quality of RCT designs themselves may be a secondary outcome measured in such evaluations to see if the CONSORT-SPI reporting guidance has a knock-on effect on the quality of the trials themselves.

Individuals' use of CONSORT-SPI should also be tracked, for example through the number of times the article is mentioned or cited in trial publications. Organisational implementation of CONSORT-SPI could be assessed via the number of journals and organisations who endorse CONSORT-SPI or CONSORT more generally yet published social and psychological intervention trials. More general support by the scientific community of reporting guidelines could be assessed by the endorsement of other reporting guidelines by organisations in the social and behavioural sciences as well as the development of reporting guidelines by influential societies and organisations.

#### **4. Discussion**

The success of a guideline depends in large part on the quality of its implementation strategy, the degree to which this strategy is developed in advance, and the ability to carry out or accomplish the strategy as planned.<sup>27</sup> A four-step method was used to identify in advance who is targeted by the CONSORT-SPI dissemination and implementation strategy, what barriers and enablers need to be addressed, how these will be addressed, and how to understand the impact of the guideline. Several theoretical frameworks were used to inform this method. The Guideline Implementability Framework was particularly helpful in identifying formatting issues of the guideline to address.<sup>24</sup> For example, though not explicitly organised this way in the checklist, the Extension Statement and the E&E will explain how the same checklist item can be applied to various types of trials, interventions, and disciplinary areas. The CFIR<sup>27</sup> was helpful in organising the different levels at which barriers and enablers might exist, particularly focusing on those related to "organisations" broadly construed, such as journals and funding organisations. The PARIHS Framework<sup>26</sup> helped to in understanding the guideline at point-of-use. The TDF<sup>12</sup> was immensely helpful in matching identified barriers and facilitators to actual

techniques and policies via its link to the Behaviour Change Wheel.<sup>12,28</sup> The prevalence of social influences as potential barriers and enablers seemed to indicate these influences as a key new area to target for reporting guideline dissemination and implementation.<sup>12</sup>

#### ***4.1 Strengths and Limitations***

Compared to previous reporting guideline projects, this dissemination and implementation strategy gains strength from its explicit grounding in theoretical frameworks. However, the methodology employed involved selective inclusion of frameworks rather than a full systematic review that identified frameworks according to strict eligibility criteria. It is worth noting, however, that these frameworks are among the most prominent in the field of implementation science, with the TDF recently having a special section in *Implementation Science*, indicating the utility of continuing work on this popular framework.<sup>9</sup> In addition to the selection of frameworks to use, it was also the sole judgment of the DPhil candidate as to which theoretical domains within these frameworks were relevant, how to interpret them in the context of CONSORT-SPI and reporting guidelines, and which implementation intervention techniques or policies are best to address them. It is entirely possible that other domains were relevant, that the proposed relevant domains were somewhat misconstrued, and other implementation intervention components may be helpful. This chapter aimed to improve on what has come before on dissemination and implementation plans specific to particular reporting guidelines, which at the time of writing is quite minute, by both explicating a coordinated strategy and trying to base it on existing theories given that little empirical evidence exists for implementation strategies outside of journal endorsement.

Another limitation to note of the proposed strategy is that it would require many resources and personnel to carry out in full. For example, a participant at the CONSORT-

SPI consensus meeting noted that, at the consensus meeting for CONSORT for Non-Pharmacologic Treatments,<sup>33</sup> the idea was also proposed to develop different versions of the E&E tailored to specific disciplines. However, only the generic version was published given the resources needed to develop tailored versions, and the belief that a generic version would be sufficient and is quite time-consuming to produce on its own. Though the CONSORT-SPI project has some resources to begin this strategy, the voluntary contribution of members of the International Advisory Group, consensus meeting participant group, and other stakeholders would be needed. The scarce time that these collaborators could offer needs to be used strategically as a result. As such, it is essential that each person's activities in the dissemination and implementation strategy are coordinated rather than at cross-purposes.

#### ***4.2 Conclusions***

This chapter provides explicit and theoretical consideration of barriers and enablers to use and implementation of CONSORT-SPI. These considerations not only have helped to generate new techniques to implement reporting guidelines, but they help old techniques to be seen in a new light as part of a theoretically-coordinated strategy. To further inform this strategy, an online forum, such as a Wiki page, should be used to facilitate evaluation of guideline impact as well as a mechanisms for feedback from authors, editors, reviewers, and funders that will assist with continual development of CONSORT-SPI. Such a forum could be used to discuss new, relevant evidence related to social and psychological intervention trials, and to asking the wider scientific community to provide feedback on their experiences of using the guideline.<sup>20</sup>

A coordinated dissemination and implementation plan that also evaluates the impact of the guideline can help the project team know if CONSORT-SPI is working as intended.

It can also help the reporting guideline field better understand how reporting guidelines are used and how to better increase the chances of their impact beyond strategies of journal endorsement. Evaluation data can also be used as audit and feedback to further help individuals change reporting practices and organisations change reporting standards. To assist such monitoring and evaluation of guideline use, guideline developers should provide journals with performance measures to audit their compliance with the checklists.<sup>24</sup> If such activities are coordinated with the EQUATOR Network, the CONSORT-SPI project could help further bring the social and behavioural sciences on board to reporting guidelines and standards in general.

## References

1. Simera I, Moher D, Hoey J, Schulz KF, Altman DG. A catalogue of reporting guidelines for health research. *European Journal of Clinical Investigation*. 2010;40:35-53.
2. Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR Network's Survey of Guideline Authors. *PLoS Medicine*. 2008;5:869-874.
3. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
4. Moher D, Weeks L, Ocampo M, et al. Describing reporting guidelines for health research: A systematic review. *Journal of Clinical Epidemiology*. 2011;64(7):718-742.
5. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *MJA*. 2006;185:263-267.
6. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews*. 2012;11:MR000030.
7. Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: Interrupted time series analysis. *BMJ*. 2012;344:e4178.
8. Eccles MP, Armstrong D, Baker R, et al. An implementation research agenda. *Implementation Science*. 2009;4:1-7.

9. Francis JJ, O'Connor D, Curran J. Theories of behaviour change synthesised into a set of theoretical groupings: Introducing a thematic series on the theoretical domains framework. *Implementation Science*. 2012;7:35.
10. French SD, Green SE, O'Connor DA, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: A systematic approach using the Theoretical Domains Framework. *Implementation Science*. 2012;7:38.
11. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. *Implementation Science*. 2012;7:50.
12. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Science*. 2012;7:37.
13. Altman DG. Endorsement of the CONSORT statement by high impact medical journals: A survey of instructions for authors. *BMJ*. 2005;330:1056-1057.
14. Smidt N, Rutjes AW, van der Windt DA, et al. The quality of diagnostic accuracy studies since the STARD statement: Has it improved? *Neurology*. 2006;67(5):792-797.
15. Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT statement by high impact factor medical journals: A survey of journal editors and journal 'Instructions to Authors'. *Trials*. 2008;9:20.
16. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.

17. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998;2(3).
18. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*. 2008;148(4):W60-W66.
19. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698-702.
20. Michie S, Abraham C, Eccles MP, Francis JJ, Hardeman W, Johnston M. Strengthening evaluation and implementation by specifying components of behaviour change interventions: A study protocol. *Implementation Science*. 2011;6:10.
21. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qualitative Health Research*. 2005;15:1277-1288.
22. Teti M, Bowleg L, Cole R, et al. A mixed methods evaluation of the effect of the Protect and Respect Intervention on the condom use and disclosure practices of women living with HIV/AIDS. *AIDS & Behavior*. 2010;14:567-579.
23. Corbin J, Strauss AC. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London, UK: Sage Publications; 2009.
24. Gagliardi AR, Brouwers M, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. *Implementation Science*. 2011;6:26.

25. Gagliardi AR, Brouwers M, Bhattacharyya OK. The guideline implementability research and application network (GIRAnet): An international collaborative to support knowledge exchange: Study protocol. *Implementation Science*. 2012;7:26.
26. Stetler CB, Damschroder LJ, Helfrich CD, Hagedorn HJ. A guide for applying a revised version of the PARIHS framework for implementation. *Implementation Science*. 2011;6:99.
27. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*. 2009;4:50.
28. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*. 2011;6:42.
29. Collins FS, Tabak LA. NIH plans to enhance reproducibility. *Nature*. 2014;505:612-613.
30. Michie S, Richardson M, Johnston M, et al. The Behavior Change Technique Taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*. 2013;46:81–95.
31. APA JARS Group. Reporting standards for research in psychology: Why do we need them? what might they be? *American Psychologist*. 2008;63:839-851.
32. Altman DG. Endorsement of the CONSORT statement by high impact medical journals: a survey of instructions for authors. . *BMJ*. 2005;330:1056-1057.
33. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT Group. Extending the CONSORT Statement to randomized trials of nonpharmacologic

treatment: Explanation and Elaboration. *Annals of Internal Medicine*.  
2008;148:295-309.

## Chapter 8: Conclusions about the CONSORT-SPI Project

### Overview of Thesis

This dissertation first discussed recent developments in intervention research that led to innovative uses of randomised controlled trials (RCTs) to evaluate social and psychological interventions (Chapter 1). These new developments warranted a project plan to investigate the need for a new reporting guideline for social and psychological intervention trials (Chapter 2). Two systematic reviews indicate that reporting quality of social and psychological intervention trials is poor, and that previous reporting guidelines did not sufficiently follow rigorous practices for guideline development and dissemination *while also* containing reporting standards needed for this area (Chapter 3). An online, international Delphi process with 384 stakeholders in social and psychological intervention research identified 58 items to consider for inclusion in the CONSORT-SPI checklist—including all but one of the items in the CONSORT 2010 checklist—while also providing constructively critical views about trial methodology and reporting in the social and behavioural sciences (Chapter 4). A 3-day consensus meeting of 31 stakeholders from the Delphi process led to the selection of 14 extended items for the CONSORT-SPI checklist and suggestions for the content of an Explanation and Elaboration (E&E) document (Chapter 5). Insights from the systematic reviews, Delphi process, and consensus meeting were then used along with key social and psychological intervention research literature to draft the CONSORT-SPI checklist and E&E template (Chapter 6). A coordinated dissemination and implementation strategy was then developed using lessons from the reporting guideline literature, feedback from the consensus meeting, and frameworks for implementing interventions and clinical guidelines (Chapter 7). This chapter addresses how this thesis contributes to previous knowledge and its implications for future research.

## 1. What We Already Knew

Though there are many scientific approaches for examining causal relationships, experimental methods are currently promoted for the reliable estimation of the causal effects of interventions. Specifically, experiments that use a counterfactual logic and create a comparison group by randomly assigning participants to interventions are favoured methods for controlling potential confounding factors that could limit causal inferences.<sup>1,2</sup> The increased use of RCTs to evaluate interventions has led to increased attention to how they are both conducted and reported.<sup>3</sup> RCT reports need to be sufficiently transparent in order to facilitate critical appraisal, which is vital to interpret the legitimacy and relevance of RCT findings to particular “real-world” contexts.<sup>4</sup> Incomplete and unclear reporting hinders the appraisal of RCTs, leaving readers susceptible to implementing biased or irrelevant research findings in their professional settings. Likely as a result of the complexities of social and psychological intervention RCTs, there is significant evidence suggesting poor reporting of these trials in disciplines as diverse as criminology,<sup>5</sup> education,<sup>6</sup> psychology,<sup>7</sup> public health,<sup>8</sup> and social work.<sup>9</sup>

Over the last 15 to 20 years, medical researchers have made a concerted effort to improve the reporting of health research by developing explicit reporting standards that provide recommendations on the content that study reports should consistently and transparently include to allow for critical appraisal.<sup>10</sup> Reporting standards are generally collated and disseminated through reporting guidelines, which typically include a checklist, flow diagram, and explicit text to guide authors in reporting a specific type of research.<sup>11</sup> The most well-known, influential, and widely used reporting guideline for RCTs is the Consolidated Standards of Reporting Trials (CONSORT) Statement, which assists authors in clearly and sufficiently reporting essential information about RCTs.<sup>11,12</sup> Since its original publication in 1996, the CONSORT Statement has increased awareness

of issues surrounding proper reporting of research, setting a higher industry standard.<sup>11-17</sup> Though other CONSORT Extensions exist,<sup>18-20</sup> many researchers have proposed that social and psychological intervention trials may benefit from tailored CONSORT reporting standards specific to their areas.<sup>6,21,22</sup>

## **2. What This Thesis Adds**

### ***2.1 Systematic Reviews***

A two-part study involved (1) a systematic review of reporting guidelines and reporting quality assessment tools related to social and psychological intervention trials, and (2) a review of the reporting quality of social and psychological intervention trials. The first review yielded 14 reporting guidelines and 5 reporting quality assessment tools with 147 non-redundant reporting standards in total. Compared to other documents in the medical, social, and behavioural sciences, CONSORT guidelines were more likely to use recommended best-practices for preliminary planning of producing the guideline, consensus development techniques for producing the guideline, and publishing and disseminating the guideline for greater uptake. In addition, CONSORT guidelines tend to have considerably higher citations per year than other guidelines in medicine and the wider behavioural and social sciences. While no current guidance was specifically developed for or sufficiently incorporated all key standards for social and psychological intervention trials, 89 distinct reporting standards were identified that were either not currently in CONSORT guidelines or were tailored versions of CONSORT items for intervention trials in the social and behavioural sciences. In summary, existing reporting guidelines have important limitations in their development, dissemination, and content as they relate to social and psychological intervention RCTs.

A hand search of the 2010 Table of Contents of the 10 highest impact factor

journals each in clinical psychology, criminology, education, public health, and social work (50 journals total) yielded 309 RCTs of social and psychological interventions. The number of eligible trials published that year ranged from 1 to 39 per journal, with a median of 3 trials per journal. Only 14 of the 50 journals (28%) referenced a reporting guideline in their “Instructions to Authors” section, and an additional 2 journals (4%) provided some advisory text about reporting certain aspects of intervention studies. Only 9 journals (18%) required trials to be registered in a trial registry prior to publication and to report this information. Based on the 147 reporting standards identified in the first systematic review in this study, reporting quality of these trials was poor overall, and it did not vary greatly by discipline. In short, important details are routinely missing from social and psychological RCT publications, and many leading journals in the social and behavioural sciences do not ask authors to follow reporting standards.

## ***2.2 Delphi Process***

An online, international Delphi process was conducted to identify those areas in the reporting of social and psychological intervention trials that should be considered for inclusion in the CONSORT-SPI checklist. A total of 384 researchers, journal editors, research funders, practitioners, policy-makers, and consumer group representatives from 32 countries were recruited through a multi-step, iterative approach involving snowball sampling.<sup>23,24</sup> In total, 58 items were recommended for inclusion in the CONSORT-SPI checklist. All but one of the items from the CONSORT 2010 Statement—Item 23 on trial registration—reached consensus for inclusion in the CONSORT-SPI guidelines during Round 1. The other items recommended for inclusion varied in their focus and their section of a trial manuscript. In summary, stakeholders in social and psychological intervention research support adherence to items in the CONSORT 2010 Statement when reporting

RCTs in this area, yet they also desire additional items currently absent from the CONSORT checklist that are tailored to key aspects of these RCTs, such as intervention theory of change, implementation, trial context, and outcome measurement.

This Delphi process provided helpful and rich information for the CONSORT-SPI consensus meeting and guideline document write-up. Across comments in both rounds of the Delphi process, participants largely highlighted the desire for a minimal, user-friendly checklist, as many were concerned that the 36 items recommended for the checklist in Round 1—let alone the 58 items in total—were too many. Many participants also noted that, given the diversity of disciplines and intervention techniques targeted by this guideline, the checklist should be worded as clearly as possible, and the guidance as a whole needs to apply to all possible social and psychological intervention trials. Specifically, participants wanted the guidance to be sensitive to the fact that some interventions targeted individuals, others groups of individuals, others geographic units or places, and some a mixture of some or all of the above. In addition, trials may be at certain “phases”—such as the pilot, efficacy, or effectiveness phase—so guidance recommendations need to be nuanced with regards to the maturity of the intervention.

### ***2.3 Consensus Meeting***

A three-day consensus meeting was held in March 2014 in Oxford, UK with 31 participants from the Delphi process in order to select the content of the CONSORT-SPI checklist, discuss guidance to include in the CONSORT-SPI E&E document, and consider how best to disseminate and implement guideline documents. In total, 14 additional items were proposed for the CONSORT-SPI checklist across all sections of the manuscript. Many items relate to information typically provided in the introduction and methods sections. A narrative analysis of the discussions at the consensus meeting offered a record

of the process by which checklist items were selected and which content should go into the explanation of items in the E&E. Participants also noted several issues that they thought needed to be highlighted at the beginning of the CONSORT-SPI E&E.

#### ***2.4 E&E Document***

A template for the CONSORT-SPI E&E document details the rationale for each item, the principles and evidence underlying it, and examples of how to report in adherence to each item. To develop the template, examples of reporting according to the checklist were found for each item from the database of social and psychological intervention RCTs from the systematic review in Chapter 3 of this thesis, as well as other prominent trials in this area. The rationale for each reporting item is based on feedback from the CONSORT-SPI Delphi process and the consensus meeting, as well as guidance from previous reporting guidelines in the social and behavioural sciences. Compared to other CONSORT Extension E&E documents, the CONSORT-SPI E&E will provide guidance for every item—not just the extended items—in light of the early stages of using reporting guidelines and CONSORT in this area. A number of changes to the CONSORT-SPI checklist were offered as a first attempt to address concerns about some aspects of CONSORT 2010 that need adapting for social and psychological interventions. However, the final version of both the checklist and the E&E will likely involve many revisions of the template in this dissertation, and any proposed changes to CONSORT 2010 items in the CONSORT-SPI checklist will ultimately need approval from the CONSORT Group.

#### ***2.5 Dissemination and Implementation Strategy***

A coordinated dissemination and implementation strategy for CONSORT-SPI was developed via a systematic approach. The first step involved clearly specifying which

stakeholders need to do what differently in order for CONSORT-SPI to be used and for the reporting quality of social and psychological intervention RCTs to be improved. These include individuals intended to use CONSORT-SPI (e.g., authors, peer-reviewers), organisations intended to implement CONSORT-SPI (e.g., journals, funding organisations), and members of the wider social and behavioural science community intended to support reporting guidelines (e.g., research societies, publishing companies). Several theoretical frameworks from the implementation science and clinical guideline literatures were then used to identify potential barriers and enablers of use and implementation of CONSORT-SPI. Modifiable barriers and enablers of guideline use were identified across guideline-, individual-, organisational-, and system-levels. Implementation techniques and policies were identified from previous guidance and evidence related to disseminating completed reporting guidelines, the initial impact strategy for CONSORT-SPI, and an analysis of discussions at the CONSORT-SPI consensus meeting. The Behaviour Change Wheel<sup>25</sup> was used to map these techniques and policies to identified barriers and enablers. The final step consisted of a plan to measure and understand desired behaviour change. If enacted, this plan will hopefully improve the actual uptake and use of CONSORT-SPI guidance, provide a model for other reporting guideline teams to follow when constructing a dissemination and implementation strategy, and lead to future empirical research on the best ways of implementing reporting guidelines.

### **3. Strengths and Limitations of this Thesis**

Several strengths and limitations should be noted about this thesis, both within and across its phases. Firstly, awareness of reporting standards and guidelines has increased in the time between this project's analysis of reporting quality (Chapter 3) and the

development of CONSORT-SPI. As such, it is possible that some improvements in reporting have occurred or are currently underway compared to the estimates in the systematic review.<sup>26</sup> Moving forward, this project would benefit from a re-assessment of the state of the social and psychological intervention literature—though likely less rigorous than the systematic review, due to resource constraints—to best integrate this new guidance within social and behavioural science publication practices.

Another topic to comment on relates to the disciplines explicitly targeted by the project. A delineated set of disciplines were targeted from the outset of the project to help ensure that core constituencies were involved in guideline development and subsequent uptake. However, there are no doubt other disciplines involved in social and psychological intervention research that have not been as deliberately involved in this project as criminology, education, psychology, public health, and social work. Whereas the Delphi process recruited researchers identifying with well over a dozen disciplines, the consensus meeting was limited in number. As the items included in the checklist are minimal, it is possible that the guidance may not address all areas desired by other stakeholders not as heavily recruited for this project. To promote uptake by stakeholders from other areas, such as economics or international development, the Project Executive may need to consider directed outreach to these stakeholders in dissemination activities, such as editorials or potentially further tailored E&E documents.

Several aspects of the Delphi process are worth noting. The Delphi process in this project was demonstrably useful in engaging a large number of stakeholders internationally and across disciplines. It was also helpful in providing some prioritisation of potential checklist items for consensus meeting participants—although it did not clarify the relative importance of items for the checklist to the degree hoped for at the outset. Despite these benefits, as a method of stakeholder consultation, the Delphi process did not technically

provide empirical support for items in terms of risk of bias or importance to generalisability, but rather aggregated opinions useful for decision-making. As such, the reliability of deciding checklist content primarily by the Delphi process results would be questionable. Consequently, the consensus meeting took the approach of using Delphi results as stakeholder advice rather than as hard and fast decisions upon which to base checklist items. This arrangement is similar to research projects that combine stakeholder consultation like the Delphi process—seeking others’ opinions though not necessarily acting on them—with stakeholder collaboration like the consensus meeting—an ongoing partnership in which decision-making is shared.<sup>27</sup> With this understanding of the Delphi process, other methods for stakeholder consultation may be as or more useful than modified Delphi approaches for developing reporting guidelines such as CONSORT-SPI. Examples to explore in future projects include the ExpertLens online iterative system<sup>28</sup> or the RAND/UCLA Appropriateness Method.<sup>29</sup>

Related to the above concern is the primary sampling of participants geographically from the US and UK, and disciplinarily from those with more “psychological” than “sociological” backgrounds. In absolute numbers, the diversity of engaged stakeholders was large compared to other guideline projects, and ensured that perspectives from these groups were included in the guideline development process. However, in relative terms, those outside of the US and UK or who didn’t have a training in disciplines related to psychology were a smaller proportion of participants in this project. These sample characteristics were partly due to take-up rates in the Delphi process and consensus meeting by those from these backgrounds, as well as a purposive sampling because these geographic regions and disciplines produce large amounts of the trials targeted by CONSORT-SPI. Moreover, the views of those from stakeholder groups with less representation were particularly noted during the consensus meeting to help ensure that

these “minority” groups will ultimately obtain guidance of use to them. Nonetheless, the inclusion of more people at the consensus meeting who *strictly* do policy interventions, for example, would have been beneficial so that such participants could have had examples in mind and present them to the group during discussion and voting. Actions moving forward could seek to include all stakeholders who fit the definition of social and psychological interventions based on mechanisms, particularly those not as engaged to date. The views of such stakeholders should be especially collated for any future updates to the guidelines.

One issue flagged from the outset of the process was the need for a guideline like CONSORT-SPI, given the existence of the CONSORT-NPT extension.<sup>18</sup> With a draft CONSORT-SPI checklist, it should be noted that many of the additional items in CONSORT-SPI are similar to CONSORT-NPT in concept, such as details of the intervention, as well as eligibility criteria for settings and providers. However, for these similar items, there is distinctive language, theories, examples, and discussed disciplines in the CONSORT-SPI guidance, which is intended to lead to greater uptake of CONSORT guidance by those in the social and behavioural sciences widely construed. Moreover, many items in the CONSORT-SPI checklist are not found in CONSORT-NPT, such as how the intervention is hypothesised to work, access to materials on delivering the intervention, the role of the intervention developer in the trial, and incentives offered during the trial. Some of these items may be missing in CONSORT-NPT given its publication in 2008 and the developments in intervention research since then. The planned update of the CONSORT-NPT guidance will indicate whether this is the case, particularly as materials from the CONSORT-SPI project have been shared with the CONSORT-NPT group. Regardless, these items were deemed essential for social and psychological intervention trials by the CONSORT-SPI consensus meeting participants and need to be collated in their own checklist for this particular area of intervention research.

For the project as opposed to the thesis, the development of the E&E will prove a significant task. The draft template developed thus far provides a crucial starting point. Its content synthesises insights from the early phases of the project, including from previous reporting guidelines, the Delphi process, and the consensus meeting. As a result of using these sources to develop the template, some sections are quite substantial (such as the section on interventions), whereas others are currently minimal (such as sections on stopping rules). The size of sections reflects the amount of attention given to topics within previous guidelines and the consensus methods of this project, which likely reflect their importance in this area to some degree. However, for the final E&E publications, some sections may need to be more concise while others may need to be bolstered by evidence and theory from the wider intervention literature. Such changes are planned for via refinement of the template by the Project Executive, and adaptation of the document to particular disciplines by consensus meeting participants—who will likely provide other examples of good reporting than those currently serving as “proof of concept” in the template. This process is expected to take significant time and resources, particularly as piloting the E&E and checklist would be helpful to do prior to publication.

Dissemination and implementation have been important aspects of this project from the outset, and the coordinated strategy reported in Chapter 7 should facilitate knowledge translation moving forward. However, many of the suggested techniques and policies within this implementation strategy lack empirical evidence and thus mostly rest on conceptual and theoretical support alone. Better empirical evidence for the proposed barriers, enablers, techniques and policies are needed, such as the accumulating evidence for introducing reporting guidelines into editorial processes.<sup>15</sup> Future research should seek to further build such an empirical evidence base for knowledge translation strategies as applied to the reporting guideline field, and the CONSORT group is currently undertaking

further explorations in this area.<sup>15</sup> One potential avenue of research is to develop implementation interventions using the Behavior Change Technique (BCT) Taxonomy.<sup>30</sup> Based on expert consensus, the BCT Taxonomy provides a list of 93 techniques—which are hierarchically classified in 16 clusters—that are used in behaviour change interventions. Activities within dissemination and implementation strategies for reporting guidelines could be defined according to this taxonomy in order to specify the actual content of interventions intended to change behavior, as well as identify appropriate techniques to address the barriers and enablers to guideline use.

A particular strength of this thesis as it relates to the project overall is the transparency it provides about the phased process of investigating the need for, developing, and disseminating CONSORT-SPI. Examples of such transparency include a full project protocol, detailed information about preliminary systematic reviews, detailed thematic analysis of Delphi responses, narrative analysis and comprehensive detail about the consensus meeting and the votes for checklist items, and a replicable method and template for disseminating and implementing reporting guidelines. Of highlight are the details about the consensus meeting discussions and votes via clickers, as well as a record of the evolution of draft checklists from before the Delphi process to after the consensus meeting (Appendices M through R) to the most up-to-date draft at the time of submission (Table 2 in Chapter 6). In addition to facilitating critical appraisal of this thesis and the studies within it, such transparent reporting should also provide valuable insights for future reporting guideline development teams.

#### **4. Implications for Future Research**

The CONSORT-SPI guidance as it stands now is still a work in progress, and to some degree it will always be a living document that will require improvements and

updates as the area develops. CONSORT-SPI and other reporting guidelines are only tools, and fallible ones at that; they do not replace the need for prudence and good judgment by those using them. As such, future research needs to be conducted on the impact of the guideline as well as needs to update its content over time. The success of this project—both in terms of its more immediate use as well as its longevity—depends on widespread involvement and agreement among key international stakeholders in research, policy, and practice. Many journal editors have already agreed to participate, and the Project Executive hopes that other researchers and stakeholders will volunteer their time and expertise to help disseminate as well as provide constructively critical feedback about its utility to the area. These stakeholders could also assist in assessing improvements in reporting as well as the success of dissemination and implementation according to the strategies proposed in this thesis.

#### ***4.1 Future Research for Social and Psychological Intervention Researchers***

Feedback during the Delphi process and consensus meeting have highlighted numerous areas that would benefit from more investigation by social and psychological intervention researchers, which may subsequently prompt updates to the CONSORT-SPI checklist and E&E. Firstly, future research should look to develop greater empirical evidence on essential features of context and aspects of intervention implementation. Many participants in this project noted that details about context and intervention implementation are important to report in manuscripts on social and psychological intervention RCTs. Beyond that, there was considerable disagreement regarding what aspects of context and implementation required reporting. For example, participants in the Delphi process simultaneously (and strongly) argued for and against the qualifications of intervention providers as crucial information. Others noted that definitions of context can vary quite

widely, and specifying which aspects were “important” is beyond the scope of current knowledge. As context and implementation are two areas of social and psychological intervention research that particularly spawned the motivation for this project, the field would benefit from future research that seeks to develop empirical knowledge about these topics.

Another area for future research is the nature of harms and unintended consequences in social and psychological interventions. These were rarely reported in trials examined in the systematic review in Chapter 3, and many Delphi participants noted that the field in general does not give much attention to detecting—let alone reporting—this information. Recent evidence in the area of psychological interventions provides empirical support to this view.<sup>31</sup> Moving forward, methodologists should provide guidance on how to consider the nature of harms in this context, and trialists should actively theorise about and aim to detect potential harms in their trials. Upcoming guidance for process evaluations in complex public health intervention trials should prove very useful in this regard.<sup>32</sup>

One final area to note is trial registration. Practically no trials in the systematic review in Chapter 3 reported trial registration numbers, and many participants in the Delphi process expressed views that the area of social and psychological intervention was not yet ready for trial registration nor did it have appropriate registries. In fact, the CONSORT 2010 item on trial registration was the only item from the CONSORT 2010 checklist that did not get recommended in the Delphi process for inclusion in the CONSORT-SPI checklist. Recent editorials have sought to combat these views and provide guidance on how to register trials.<sup>33</sup> Future outreach should be conducted to ensure that trial registration becomes a norm in this area, as it is a crucial practice for high-quality trials and an accurate scientific record of them.

#### ***4.2 Other Reporting Guidelines***

The studies within this thesis and the insights gained therein suggest that at least two new reporting guideline projects would be useful for the field. Firstly, many participants in the Delphi process indicated that the reporting standards being proposed would be equally relevant and useful to non-randomised evaluations of social and psychological interventions. The developers of the Transparent Reporting of Evaluations with Non-Randomized Designs (TREND) Statement noted in its publication<sup>34</sup> that they did not include a large diversity of stakeholders in the development of TREND. They also indicated that TREND could benefit from tailoring to particular disciplines as well as specific non-randomised designs, given that each design may lead to quite substantial differences in what needs to be reported. As such, TREND—which was published over a decade ago now—could benefit from an update, in particular if building on the insights and professional networks from the CONSORT-SPI project.

Secondly, in planning and conducting the Delphi process, the DPhil candidate systematically retrieved and reviewed all previous reporting guideline projects that utilised the Delphi method. In doing so, it was discovered that many of these publications gave very little details about the methods used, hindering their utility as pedagogical tools. Poor reporting of Delphi processes also obstructed critical appraisal of the rigour of these processes, which underpin the content of reporting guidelines that are meant to influence entire areas of scientific research. Leading experts in reporting guidelines recommend using the Delphi method for the development of any new reporting guideline,<sup>11</sup> yet guidance has been lacking on exactly how to design these studies and what to report. Researchers attempting to develop core outcome sets for intervention research<sup>35</sup> and to identify healthcare quality indicators<sup>36</sup> have identified deficiencies in reporting Delphi processes in these areas; they along with others<sup>37</sup> have constructed recommended

checklists of items that should be reported in studies using the Delphi technique. A set of consolidated standards for reporting Delphi processes would be useful to improve the quality of reporting these studies. Such a guideline could have an immense impact, as this method is not only used to develop reporting guidelines, but other very important outputs in health research as well, such as practice and policy guidelines, core outcome sets, research priorities within disciplines, curriculum for health-profession qualifications, and health indicators.

## 5. Closing

RCTs are reliable *only* when they are designed and executed properly. An RCT, in itself, is not a guarantee of internal validity.<sup>38</sup> The inappropriate use of RCTs and the presence of biases can lead to inaccurate data. A significant concern is that, given the current reputable status of RCTs, some research consumers may uncritically accept misleading trial results and could subsequently believe interventions to be more effective, ineffective, or harmful than they actually are.<sup>39</sup> If professionals informed by biased trials subsequently come to false conclusions about the effectiveness of various interventions, then the purpose of using research evidence to inform policy and practice decision-making is undermined.<sup>21</sup> As such, transparent, usable reports are needed to achieve the expected returns on investment of RCTs. When RCT reports cannot be used, resources are wasted, and stakeholders do not obtain the anticipated benefits from their investment in the research community.

Reporting standards offer a partial solution to intervention trials that are missing critical information for using research. Reporting guidelines that codify these standards aim to help the research community use its scarce resources efficiently, effectively, and ethically. The goal of the CONSORT-SPI guideline in particular is to increase the quality

of reports of social and psychological intervention RCTs. As the volume of social and psychological intervention literature is constantly growing, transparent reporting of these RCTs is especially crucial for synthesising the evidence on interventions to create a cumulative body of scientific knowledge.<sup>34</sup> Use of these guidelines by authors, editors, peer-reviewers, and funders should hopefully lead to higher-quality research to inform policy and practice decision-making. This guideline is just one step toward improving reports of many designs for evaluating social and psychological interventions, which will hopefully be launched by this project and addressed by subsequent ones. Moving forward, stakeholders from disciplines that frequently research these interventions are invited to join this important effort.

## References

1. Brady HE. Causation and explanation in social science. In: Coolier D, Brady HE, Box-Steffensmeier JM, eds. *The Oxford handbook of political methodology*. Oxford: Oxford University Press; 2008.
2. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth; 2002.
3. Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *Journal of Clinical Epidemiology*. 2011;64:136-144.
4. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology*. 2004;4:22.
5. Perry AE, Weisburd D, Hewitt C. Are criminologists describing randomized controlled trials in ways that allow us to assess them? Findings from a sample of crime and justice trials. *Journal of Experimental Criminology*. 2010;6:245-262.
6. Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. A comparison of RCTs in health and education. *British Educational Research Journal*. 2005;31(6):761-785.
7. Stinson JN, McGrath PJ, Yamada JT. Clinical trials in the Journal of Pediatric Psychology: Applying the CONSORT statement. *Journal of Pediatric Psychology*. 2003;28 159-167.
8. Semaan S, Kay L, Strouse D, et al. A profile of U.S.-based trials of behavioral and social interventions for HIV risk reduction. *Journal of Acquired Immune Deficiency Syndromes*. 2002;30:S30-S50.

9. Naleppa MJ, Cagle JG. Treatment fidelity in social work intervention research: A review of published studies. *Research on Social Work Practice*. 2010;20:674-681.
10. American Educational Research Association. Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*. 2006;35(6):33-40.
11. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010;7(2):e1000217.
12. Moher D, Jones A, Lepage L, for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA*. 2001;285:1992-1995.
13. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637-639.
14. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *MJA*. 2006;185:263-267.
15. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews*. 2012;11:MR000030.
16. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Controlled Clinical Trials*. 2002;23:380-388.

17. Moher D, Altman DG, Schulz KF, Elbourne DR. Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *CMAJ*. 2004;171(4):349-350.
18. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT Group. Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: Explanation and Elaboration. *Annals of Internal Medicine*. 2008;148:295-309.
19. Gagnier J, Boon H, Rochon P, Barnes J, Moher D, Bombardier C. Reporting random controlled trials of herbal medicines. *Explore: The Journal of Science and Healing*. 2006;2(2):143-149.
20. MacPherson H, White A, Cummings M, et al. Standards for Reporting Interventions in Controlled Trials of Acupuncture: The STRICTA recommendations. *Journal of Alternative & Complementary Medicine*. 2002;8(1):85-89.
21. Boutron I, John P, Torgerson DJ. Reporting methodological items in randomized experiments in political science. *The ANNALS of the American Academy of Political and Social Science*. 2010;628(112):112-131.
22. McGrath PJ, Stinson J, Davidson K. Commentary: The Journal of Pediatric Psychology should adopt the CONSORT statement as a way of improving the evidence base in pediatric psychology. *Journal of Pediatric Psychology*. 2003;28:169-171.
23. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*. 2000;32(4):1008-1015.
24. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311(376).

25. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*. 2011;6:42.
26. Grant SP, Mayo-Wilson E, Melendez-Torres GJ, Montgomery P. Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PloS One*. 2013;8(5):e65442.
27. Gough D, Oliver S, Thomas J. *An introduction to systematic reviews*. London: Sage Publications; 2012.
28. Khodyakov D, Hempel S, Rubenstein L, et al. Conducting online expert panels: A feasibility and experimental replicability study. *BMC Medical Research Methodology*. 2011;11:174.
29. Fitch K, Bernstein SJ, Aguilar MD, et al. *The RAND/UCLA Appropriateness Method User's Manual*. Santa Monica, CA: RAND; 2001.
30. Michie S, Richardson M, Johnston M, et al. The Behavior Change Technique Taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*. 2013;46:81–95.
31. Jonsson U, Alaie I, Parling T, Arnberg FK. Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. *Contemporary Clinical Trials*. 2014;38:1-8.
32. Moore G, Audrey S, Barker M, et al. Process evaluation in complex public health intervention studies: The need for guidance. *Journal of Epidemiology & Community Health*. 2013;68(2):101-102.

33. Harrison BA, Mayo-Wilson E. Trial registration: Understanding and preventing reporting bias in social work research. *Research on Social Work Practice*. 2014;24(3):372-376.
34. Des Jarlais DC, Lyles C, Crepaz N, the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*. 2004;94(3):361-366.
35. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*. 2011;8(1):e1000393.
36. Boukdedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *PLoS One*. 2011;6(6):e20476.
37. Moore L, Graham A, Diamond I. On the feasibility of conducting randomised trials in education: Case study of a sex education intervention. *British Educational Research Journal*. 2003;29(5):673-689.
38. Gill CE. Missing links: How descriptive validity impacts the policy relevance of randomized controlled trials in criminology. *Journal of Experimental Criminology*. 2011;7(3):201-224.
39. Moher D, Hopewell S, Schultz KF, Altman DG. Resources for authors of reports of randomized trials: Harnessing the wisdom of authors, editors, and readers. *Trials*. 2011;12:98.