

COHORT STUDY TO ASSESS THE QUALITY AND METHODOLOGICAL CONDUCT OF CLINICAL TRIALS

by

AYODELE ODUTAYO
BHSc MD MSc (Oxon)

Brasenose College

Thesis presented for the degree of Doctor of Philosophy in Musculoskeletal Sciences

Centre for Statistics in Medicine
Botnar Research Centre
Windmill Road
Oxford
United Kingdom
OX3 7LD

Supervisors: Prof. Douglas G Altman and Assoc. Prof Sally Hopewell

Michaelmas Term 2016

ACKNOWLEDGEMENTS

There are many people who deserve thanks for enabling me to complete the research submitted for my DPhil. First and foremost, I would like to thank God for his blessing. I feel exceptionally blessed to have had the opportunity to study at the University of Oxford. I would also like to thank the Rhodes Trust and the Clarendon Trust. Both institutions provided full funding for me to study at the University of Oxford and to focus solely on learning.

I would also like to thank my co-supervisors, Prof Douglas G. Altman and Assoc. Prof. Sally Hopewell. You have both been exceptionally supportive of my doctorate and very generous with your time. You have also been encouraging when I sought to expand my learning into other areas of epidemiology and have pushed me to become self-directed in my learning. Finally, you have been very considerate of my personal well-being and allowed me to manage my doctoral training along with other important life commitments. I am forever grateful.

I would also like to thank my long standing mentor, Dr. Michelle Hladunewich. You have been one of my biggest supporters throughout my undergraduate and medical training. Working with you was a pivotal moment in my education and I would not be where I am now without the opportunities you provided to me.

To my mother and my sister, thank you both for your ongoing support of me over the last few years. I know that being away from Canada has not been easy on our family but I am grateful that you were both understanding and helpful on every occasion.

Finally, to my wife, my long term partner and my best friend, Rashida Odutayo: words cannot express how grateful I am to have you in my life. We have been on a long journey together and have spent many miles apart. However, your support through each phase has been unwavering. You love and support was a source of strength for me and allowed me to face all of the challenges of my university education. Thank you for everything that you do.

ABSTRACT

COHORT STUDY TO ASSESS THE QUALITY AND METHODOLOGICAL CONDUCT OF CLINICAL TRIALS

AYODELE ODUTAYO
BHSc MD MSc (Oxon)

Brasenose College

Thesis presented for the degree of Doctor of Philosophy in Musculoskeletal Sciences

Michaelmas Term 2016

Clinical trials should be based on well-constructed study protocols. In practice, however, trial protocols are variable in their content and quality. This thesis examined the design and planned methodological conduct of contemporary clinical trials using a cohort of clinical trial protocols. The DPhil candidate was granted access to 259 clinical trial protocols approved by the United Kingdom National Research Ethics Service in 2012. The cohort included non-randomized clinical trials as well as randomized controlled trials (RCTs) and there was no restriction by intervention, funding source or medical specialty. There were five broad findings. First, clinical trial protocols were variable in the quality of reporting of multiple methodological items, including items that have been associated with bias. Second, the majority of trial protocols included patient reported outcome measures (PROMs) but few discussed the validity of the PROM, the strategy for dealing with missing data and the sample size calculation. Third, the primary and secondary outcomes in the trial protocols were compared to the trial registry entry if available. In this analysis, agreement in the reporting of outcomes in the trial protocol and trial registry was modest.

Notably, pharmacologic trials and multicentre trials were less likely to have discrepancies between the protocol and registry as compared to non-pharmacologic and single centre trials, respectively. Fourth, multi-arm trials made up just 8% of all clinical trial protocols. Methodological items that were important for multi-arm trials were also omitted in trial protocols. Finally, phase I trials were generally safe but recommended strategies for the identification of starting doses and the administration of the novel agents were infrequently followed. Also, phase I trials with serious adverse events were not disseminated. Taken together, this thesis has provided insight into different methodological issues in the design, conduct and dissemination of clinical trials.

TABLE OF CONTENTS

Acknowledgements 2

Abstract 4

Contents 6

List of Figures 9

List of Tables 10

Preface 15

1 Introduction 17

1.1 The Importance of Randomization 18

1.2 Assessing The Quality of Randomization 22

1.3 Other Sources of Bias Beyond Randomization 24

1.4 Other Study Characteristics Contributing to Overestimation of Treatment Effects 27

1.5 Limitations of Methodological Studies of Published Trials and the Need for Clinical Trial Protocols 29

1.6 Improving the Design of Clinical Trials and Consort 31

1.7 Improving The Design of Clinical Trials and The Standard Protocol Items: Recommendations For Interventional Trials (SPIRIT) Initiative 33

2 Methods 36

2.1 Research Questions 36

2.2 Research Ethics Approval in the United Kingdom 38

2.3 Data Sources and Study Selection Criteria 39

2.4 Challenges in the Process of Acquiring Clinical Trial Protocols 40

2.5 Data Abstraction 41

2.6 Sample Size Calculation 42

2.7 Data Analysis and Statistical Considerations 42

2.8 Inclusion of Other Study Cohorts 47

3 Overview Of The Aspire Cohort And Quality Of Reporting Of Protocols 50

3.1 Chapter Specific Methods 52

3.2 Chapter Results 56

3.3 Chapter Discussion 73

4 Assessing the Use of Patient Reported Outcomes in Protocols for Clinical Trials 78

4.1 Chapter Specific Methods 82

4.2 Chapter Results 87

4.3 Chapter Discussion 95

5 Assessing Concordance of Primary and Secondary Outcomes in Clinical Trial Protocols and Trial Registries 99

5.1 Chapter Specific Methods 103

5.2 Chapter Results 109

5.3 Chapter Discussion 130

6 Assessment of the Epidemiology and Design of Multi-Arm Trials 138

6.1 Chapter Specific Methods 144

6.2 Chapter Results 149

6.3 Chapter Discussion 164

7 Assessment of the Design, Conduct, and Dissemination of Phase I Trials 171

7.1 Chapter Specific Methods 175

7.2 Chapter Results 179

7.3 Chapter Discussion 186

8 Conclusion 192

List of References 201

Appendices 226

Appendix 1: Definition of Items in SPIRIT Checklist 226

Appendix 2: Examples of Patient Reported Outcome Measures..... 237

Appendix 3: Original Research Publications Contributing To This Thesis 238

Appendix 4: Other Notable Publications During This DPhil 239

Appendix 5: Conference Abstracts And Presentations During This DPhil 241

Appendix 6: Peer Review Responsibilities During This DPhil 243

Appendix 7: Courses Completed During This DPhil 244

LIST OF FIGURES

FIGURE 3.1: Distribution of Studies Approved by Research Ethics Committees in the United Kingdom in 2012 57

FIGURE 3.2: Identification of Included Studies for the Overall Cohort 59

FIGURE 4.1: Identification of Studies Included in the Analysis of Patient Reported Outcome Measures 88

FIGURE 4.2: Comparison of Reported and Replicated Sample Sizes 94

FIGURE 5.1: Identification of Included Studies in the Analysis of Trial Registration 110

FIGURE 5.2: Concordance in Study Outcomes Between Trial Protocols and Trial Registries 115

FIGURE 6.1: Decision Tool to Determine the Need for Type I Error Adjustment in Multi-Arm Trials 147

FIGURE 6.2: Systematic Search for Cohort Studies of Clinical Trial Protocols 151

FIGURE 6.3: Identification of Studies from the ASPIRE Cohort That Were Included in the Analysis of Multi-Arm Trials 154

FIGURE 6.4: Association Between Funding Source and Reporting of Analytical Strategies for Multi-Arm Trials 163

FIGURE 7.1: Identification of Studies in the Analysis of Phase I Versus Phase II-IV Trials 180

LIST OF TABLES

TABLE 2.1: Calculation of Risk Ratios 45

TABLE 3.1: General Characteristics of Clinical Studies at Regional Offices in the United Kingdom Based on an Assessment of Study Titles and General Application Details 58

TABLE 3.2: General Characteristics of Trials Included in the ASPIRE Cohort 61

TABLE 3.3: General Characteristics of Trials With and Without Protocols Available 62

TABLE 3.4: Reporting of SPIRIT Checklist Items in Clinical Trial Protocols 64

TABLE 3.5: Reporting of SPIRIT Checklist Items Stratified as Frequent Versus Infrequent 68

TABLE 3.6: Association Between Study Characteristics and Reporting of SPIRIT Items 70

TABLE 4.1: General Characteristics of Included Randomized Controlled Trials 89

TABLE 4.2: Inclusion of Patient Reported Outcome Measures in Protocols of Randomized Controlled Trials 91

TABLE 4.3: Methodological Items Related to the Use of Patient Reported Outcome Measures as a Primary Outcome in Clinical Trials 92

TABLE 5.1: Intraclass Correlation Coefficients and Selected Regression Models for Analysis of Discrepancies Between Trial Protocols and Trial Registries 108

TABLE 5.2: General Characteristics of Included Studies in the Analysis of Trial Registration 111

TABLE 5.3: Assessment of Discrepancies Between Trial Protocol and Trial Registry with the Individual Trial as the Unit of Analysis 113

TABLE 5.4: Assessment of Discrepancies Between Trial Protocol and Trial Registry

with the Study Outcome as the Unit of Analysis 118

TABLE 5.5: Assessment of Discrepancies Between Trial Protocol and Trial Registry with the Study Outcome as the Unit of Analysis (Primary Outcomes Only) 120

TABLE 5.6: Assessment of Discrepancies Between Trial Protocol and Trial Registry with the Study Outcome as the Unit of Analysis (Secondary Outcomes Only) 122

TABLE 5.7: Association Between Study Characteristics and Discrepancies in the Classification of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome) 125

TABLE 5.8: Association Between Study Characteristics and Discrepancies in the Reporting of the Measurement Variable of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome) 126

TABLE 5.9: Association Between Study Characteristics and Discrepancies in the Analytic Metric of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome) 127

TABLE 5.10: Association Between Study Characteristics and Discrepancies in the Time Point of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome) 128

TABLE 5.11: Reasons for Non-Registration of Clinical Trials 130

TABLE 6.1: Epidemiology of Multi-Arm Trials in the ASPIRE Cohort, Swiss Cohort, Published Studies of Clinical Trial Protocols and the ESORT Cohort 152

TABLE 6.2: General Characteristics of Multi-Arm Trials 155

TABLE 6.3: Methodological Items Related to the Design of Multi-Arm Trials 157

TABLE 6.4: Methodological Items Related to Multiple Testing in Multi-Arm Trials 159

Table 6.5: Comparisons Between Analysis Strategy in Trial Protocols and the Decision Tool for the Adjustment of Type I Error Rate in Multi-Arm Trials with Multiple Testing Planned 161

Table 6.6: Potential Items for Inclusion in SPIRIT Extension for Multi-Arm Trials 168

TABLE 7.1: General Characteristics of Phase I Trials and Phase II-IV Trials 181

TABLE 7.2: Methodological Characteristics of First-in-Human Phase I Trials 182

TABLE 7.3: Trial Registration and Intended Dissemination Strategies for Phase I Compared to Phase II-IV Trials 185

TABLE 7.4: Potential Items for Inclusion in SPIRIT Extension for Phase I Trials 189

LIST OF ABBREVIATIONS

PROMs	Patient Reported Outcome Measures
SPIRIT	Standard Protocol Items: Recommendations for Interventional Trials
DPhil	Doctor of Philosophy
UK	United Kingdom
NRES	National Research Ethics Service
CSM	Centre for Statistics in Medicine
ASPIRE	Adherence to SPIrit REcommendations
NRTs	Non-randomized clinical trials
RCTs	Randomized controlled trials
ROR	Ratio of odds ratios
CI	Confidence Interval
US	United States
FDA	Food and Drug Administration
REC	Research Ethics Committee
CONSORT	Consolidated Standards of Reporting Trials
IRAS	Integrated Research Application System
CTIMP	Clinical trial of an investigational medicinal product
CIMD	Clinical investigation or other study of a medical device
IQR	Inter-quartile interval
NIHR	National Institute for Health Research
EMA	European Medicines Agency
aRR	Adjusted Relative Risk

WHO	World Health Organization
ICMJE	International Committee of Medical Journal Editors
ISRCTN	International Standard Randomized Controlled Trial Number
EU	European Union
MABEL	Minimum Anticipated Biological Effect Level
PAD	Pharmacologically Active Dose
NOAEL	No Observed Adverse Effect Level

PREFACE

Clinical trials are considered the gold standard of research studies (1). They allow researchers to rigorously evaluate the efficacy of treatments and also allow for assessment of causal relationships between interventions and outcomes (1). However, clinical trials can be costly, time consuming and in some instances difficult to complete. Furthermore, inadequate conduct of clinical trials can be associated with bias, which is defined as "any systematic error that results in an incorrect estimate of the association between [an] exposure and [the] risk of disease" (1).

The decision to undertake a clinical trial must therefore be thoughtful, well-reasoned and based on a well-constructed study protocol. In practice however, protocols are variable in content and quality (2,3), with many providing inadequate details about items that are important to the conduct and interpretation of clinical trial results (2-4). These deficiencies may lead to improper trial conduct and the scientific literature may be skewed by spurious results. In the worst instance, treatments that are ineffective or harmful for patients may be promoted (2).

In response to the aforementioned concerns regarding clinical trial protocols, the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) Initiative was introduced in 2013 (5,6). The aim of SPIRIT is to improve the completeness of clinical trial protocols through evidence-based recommendations for a minimum set of checklist items to be addressed in protocols. The utility of the SPIRIT checklist must be evaluated and the first step in this evaluation – and herein the basis of this Doctor of Philosophy

(DPhil) thesis – is to establish a baseline measure of the quality of recent clinical trial protocols prior to the publication of the SPIRIT guideline in 2013.

Through the support of the United Kingdom (UK) National Research Ethics Service (NRES), the Centre for Statistics in Medicine (CSM) has been given unprecedented access to clinical trial protocols from research institutions in the UK. The trial protocols forming this cohort– entitled the ASPIRE (Adherence to SPIrit REcommendations) study – were approved between January and December 2012. Access to clinical trial protocols will also allow for the examination of other items related to clinical trial design and conduct.

Accordingly, this thesis is constituted of four broad sections. The first section is the introduction, which will provide an overview of the advantages of clinical trials; a description of the potential sources of bias and its impact on study findings; and a detailed description of efforts to improve the design of clinical trials and reduce bias. The second section is the methods and this section will detail the objectives of this thesis and the related research questions; provide a description of the data sources examined in this thesis; and finally, detail the analytical strategies employed in this thesis. The third section is the results, which will start with a brief background to each of the research questions, detail any specific methods use to address each research question and provide a summary of the related findings. The final section is the conclusion which will review the overall objectives of this thesis, discuss strategies that may have improved the results and discuss future areas for research. Taken together, this thesis will provide a detailed examination of different methodological issues related to clinical trial conduct and provide insights into how the design, conduct and dissemination of clinical trials can be improved.

1

INTRODUCTION

Clinical trials are prospective studies in which patients, or groups of patients, are assigned to one or more interventions to assess the effect on health outcomes. These interventions can be medications, medical devices or even behavioral or educational interventions such as smoking cessation programs. The outcomes of interest can also vary from objectively defined health outcomes such as mortality, to patient reported outcome measures (PROMs), which are reports coming directly from patients about how they feel or function.

Clinical trials have the potential to guide decision making on what interventions should be adopted into medical practice. As a result, they have become more common in the scientific literature. In a study examining trends in the publication of clinical trials, investigators searched MEDLINE for publications of multicentre clinical trials at four time points: 1995, 2000, 2005, 2010. The number of clinical trials increased from 112 in 1995 to

1273 in 2010, an increase of over 1000% (7). Specialty specific studies have also noted an increase in the number of clinical trials. For instance, a similar study demonstrated that the volume of surgical trials increased from 300 publications in 1999 to 450 publications in 2009, an increase of 50% (8).

However, not all clinical trials are equal and the design of each clinical trial should be evaluated in order to determine how the results of the trial should be interpreted. The primary design characteristic that should be considered is how the assignment to the intervention is performed. In non-randomized clinical trials (NRTs), participants are assigned to treatments based on clinical decision making (9). In randomized controlled trials (RCTs) participants are assigned using a random allocation method, whether through a random number table or computer algorithm (10). NRTs and RCTs are important sources of evidence and each is associated with specific advantages and disadvantages, as discussed below.

1.1 THE IMPORTANCE OF RANDOMISATION

NRTs are observational studies in which participants are assigned to treatments in a non-random manner. The focus of this chapter will be on prospective NRTs, wherein participants are identified, provided an intervention and prospectively followed to ascertain an outcome event. The key benefit of NRTs is that they allow for the examination of populations that may not otherwise be eligible for RCTs (9). This includes older adults and patients with comorbidities (11-15). NRTs may also allow for long term monitoring that may not otherwise be possible in RCTs.

However, the conclusions that can be drawn from NRTs may be limited because the allocation to the intervention is based on the decision making of the investigator. This can result in allocation bias or confounding, which is when the decision to assign an individual to the intervention is affected by other variables that are prognostic of the outcome of interest (9). For instance, participants with diabetes may be more likely to receive an intervention because the researchers in the NRT believe it would be beneficial. This form of decision making leads to an imbalance in the characteristics between the intervention and non-intervention group such that one group will have more participants with diabetes than the other group. If diabetes is independently prognostic for the outcome of interest – for instance, mortality – it becomes unclear whether any association between the intervention and outcome is simply due to the imbalance in the number of participants with diabetes.

Multiple strategies exist to reduce the effect of confounding, including the use of analyses stratified by the potential confounding variables, the use of multivariable regression models to account for confounding and matching participants in the study based on the confounder of interest (1). While these approaches allow for improved estimates of the association between treatments and the outcome, they can only account for potential prognostic variables that investigators are aware of and have measured (1). The contribution of unknown confounding variables will not be accounted for and may continue to bias estimates derived from the study (1).

In contrast, random allocation can readily overcome confounding because both known and unknown confounders are equally distributed in the treatment and non-

treatment groups (1). There are two components that contribute to successful randomization (16). First, a random sequence must be generated through a random number table or computer algorithm (16). Second, the random sequence must be concealed, such that the investigator recruiting participants cannot predict whether the participant will receive the treatment or the intervention (16). This latter component is known as allocation concealment and it can be achieved by ensuring that the individual generating the randomization sequence is not involved with the recruitment of participants (16). For instance, the randomization sequence can be generated and the group assignments can be provided to recruiting investigators in sequentially numbered opaque sealed envelopes. Alternatively, a centralized computer system can be used, wherein the recruiting investigator must contact an automated number to determine the treatment group to which a participant will be assigned. When random sequence generation and allocation concealment are both performed successfully, unconfounded estimates of the association between a treatment and intervention can be obtained.

Of interest, multiple studies have compared treatment effect estimates in NRTs and RCTs. In a small study, investigators examined five topics with treatment estimates obtained from a matched sample of NRTs and RCTs (17). The investigators conducted random effects meta-analyses of the estimates from the NRTs and RCTs and compared the pooled results (17). Although there were no situations in which the difference in the magnitude of the pooled treatment estimates from NRTs and RCTs was greater than what would be expected by chance (17), the use of a random effects meta-analysis warrants scrutiny. A random effects strategy weighs each study more equally and leads to wider confidence intervals in the pooled estimate, as compared to a fixed effects strategy (16).

This may have contributed to the absence of any differences between estimates from NRTs and RCTs.

In a slightly larger study, investigators compared 19 treatment estimates obtained from a matched sample of NRTs and RCTs using a fixed effects analysis (18). The authors found that in 2 of 19 cases, the summary effect estimate for NRTs was outside the 95% confidence interval (CI) of the estimate from RCTs (18). A subsequent and larger study compared 26 treatment estimates from prospectively conducted NRTs and RCTs and found that for 16 comparisons (62%), there was at least a 50% difference in the odds ratio obtained from NRTs compared to RCTs (19). Finally, a systematic review was conducted in 2014 to summarize existing methodological studies that have compared treatment effect estimates in NRTs and RCTs (20). This review analyzed data from 15 methodological studies, which contributed a mean of 178 NRTs and RCTs combined (20). Furthermore, the authors were able to include NRTs which employed more contemporary methods for adjusting for confounding, such as propensity score matching (20). Notably, this analysis found that there was little evidence of a difference in the treatment effect estimates in NRTs compared to RCTs (ratio of odds ratios [ROR] 1.08, 95% confidence interval [CI] 0.89-1.21) (20).

To date, only one study has compared the relative risks of harms in NRTs and RCTs and this analysis included 13 treatment estimates from each type of study (only 8 comparisons included absolute risks) (21). The pooled relative risks and absolute risks for NRTs and RCTs differed more than 2-fold for 7 of the 13 comparisons (54%) and 5 of 8 comparisons (62%) respectively (21). Furthermore, the absolute risk of harms in RCTs was

larger than in NRTs for 6 of 8 comparisons (75%) (21). Therefore, NRTs provided conservative estimates of the absolute risks of harms.

Accordingly, these findings suggest that while the estimates from NRTs are not discrepant from RCTs in the direction of effect, there can be differences in the magnitude of the absolute risk of harms in NRTs and RCTs. Therefore, there is still value in randomization when assigning participants to different treatments (19).

1.2 ASSESSING THE QUALITY OF RANDOMIZATION

A potential factor that was not examined in the comparisons of NRTs and RCTs was the robustness of the randomization method in the RCTs. As mentioned, randomization requires the successful generation of a random sequence and the concealment of this sequence from the investigators recruiting and assigning participants (16). If the randomization is conducted successfully, then the treatment estimates from the RCT can be considered to be unconfounded. If the randomization is not successful, then confounding may still alter the treatment estimates. Robust methods for random sequence generation include the use of a random number table or a computer based random number generator (16). Likewise, robust methods for allocation concealment include use of sequentially numbered opaque sealed envelopes, interactive voice response systems, or use of a pharmacy (16).

Using published meta-analyses of RCTs, investigators have compared estimates from studies providing a clear description of adequate randomization and allocation concealment to studies with inadequate reports. RCTs in each meta-analysis are separated into strata based on the quality of reporting in the individual published studies and a ratio of odds ratios (ROR) is calculated using logistic regression with robust standard errors (22).

Using this framework, the first study to provide empirical evidence of an association between the reported quality of allocation concealment and bias was by Schulz et al. The authors examined 250 RCTs from 33 meta-analyses (23) and found that RCTs in which the reported allocation concealment was inadequate or unclear overestimated treatment effects by at least 30% compared to RCTs in which the reported allocation concealment was adequate (23). A larger study which included 1346 RCTs in 146 meta-analyses also obtained similar results (ROR: 0.83, 95% CI 0.74-0.93; ROR <1 indicates larger effects in trials that are poorly reported) (24). Of interest, this study also stratified the trials based on whether the primary outcome was "objective" or "subjective" (24). Objective measures included all-cause mortality, measures based on laboratory tests, time based measures (e.g. pre-term birth) or surgical outcomes (24). Subjective measures included PROMs, physician assessed outcomes (including cause-specific mortality) and composite outcomes that included a subjective outcome (24). Notably, for studies with objective outcomes, there was little difference between RCTs with adequately and inadequately reported allocation concealment (ROR 0.91, 0.80-1.03) (24). However, there was a marked difference between studies that included subjective primary outcomes (ROR

0.69, 0.59-0.82), again indicating exaggerated estimates of treatment effects when adequate allocation concealment was not reported (24).

Finally, the most recent analysis by Savovic et al. involved 1973 RCTs (234 meta-analyses), and RCTs were compared based on the reporting of the randomization sequence and allocation concealment (25). The studies were also stratified based on whether the outcome was "objective" or "subjective". The authors ultimately demonstrated that trials with subjective outcomes had exaggerated treatment effects if the reporting of random sequence generation, allocation concealment or blinding was inadequate in the trial publication (25). In contrast, trials with objective outcomes did not have similarly exaggerated findings (25).

Taken together, the reported quality of randomization and allocation concealment in the published article may be associated with exaggerated treatment effects, particularly if the primary outcome of the study is subjective. When interpreting the results of NRTs and RCTs, the degree to which NRTs adjust for confounding and the quality of randomization in RCTs should be assessed. This will allow for more informed conclusions about the efficacy of interventions.

1.3 OTHER SOURCES OF BIAS BEYOND RANDOMIZATION

Beyond randomization, there are other important sources of bias stemming from the methodological conduct of NRTs and RCTs. A common source of bias is performance bias, which occurs when there are differences in the care received by the treatment and non-treatment group beyond the intervention that is being examined (9). Differences in the

type of care may be due to knowledge of the treatment the participant has received and the expectation of a specific clinical course or outcome. For instance, if the participant has received a new surgical device, they may receive more frequent follow up and intensive management compared to participants who did not receive the device. A related but different source of bias is detection bias, which occurs when the approach to the ascertainment of the outcome varies between the intervention and non-intervention group (9). For instance, if the outcome in a clinical trial is a PROM, detection bias would occur if the investigator responsible for assessing patients asks more symptom-related questions to the intervention group because they have knowledge of who received the intervention and expect a specific outcome.

Performance bias and detection bias can both be avoided through the use of blinding (9). Blinding is any strategy to make investigators, participants, or end point assessors unaware of the intervention to which the trial participant has been allocated. Blinding can be achieved through ensuring pharmacological treatments are similar in appearance taste and smell or through the use of sham procedures if the trial is examining a procedural intervention. Notably, the aforementioned meta-epidemiological study by Savovic et al. compared RCTs with blinding to RCTs without blinding and no difference was found in the ROR for studies with objective outcomes (0.93, 0.74-1.18) (25). However, the ROR for studies with subjective outcomes demonstrated that RCTs without blinding overestimated treatment effects compared to RCTs with blinding (ROR 0.78, 0.65-0.92) (25). Similar results are likely for NRTs with blinded adjudication of events compared to NRTs without blinded adjudication. However, this has not been studied.

Finally, bias may also result if a large number of participants drop out of the trial (9). Participant drop-out can occur for a variety of reasons, including a refusal from participants to continue the trial, difficulty contacting participants or a decision to exclude the participants due to non-adherence to the study requirements. However, participants who drop out after randomization may be different than participants who remain in the trial. For instance, participants may drop out due to exacerbations of their underlying disease or due to side effects from the intervention. This form of bias is known as attrition bias and can be prevented by following an intention-to-treat analysis strategy (9) wherein all randomized participants are included in the final analysis, regardless of their degree of participation in the study or their adherence to the intervention. Notably, meta-epidemiological studies have demonstrated that post-randomization exclusion of participants results in overestimated treatment effects compared to studies without post-randomization exclusions (difference in means for a continuous outcome −0.13, 95% CI −0.29 to 0.04) (26).

The importance of reducing bias should not be considered esoteric. Indeed, trials which aim to reduce bias are most crucial for clinical recommendations and decision making. For instance, investigators conducted an analysis of publicly available documents related to the approval of novel therapeutic agents by the United States (US) Food and Drug Administration (FDA) between 2005 and 2014 (27). During this time period, the FDA approved 188 novel therapeutic agents on the basis of 448 pivotal efficacy trials (27). Almost all pivotal trials were randomized (89%), double blind (80%) and had limited loss to follow up (median proportion of participants completing trials 87%) (27). Therefore, the aforementioned study characteristics represent the highest standards of methodological conduct and must be considered when interpreting the results of trials.

1.4 OTHER STUDY CHARACTERISTICS CONTRIBUTING TO OVERESTIMATION OF TREATMENT EFFECTS

Beyond biases stemming from the methodological conduct of clinical trials, other study characteristics can contribute to the overestimation of treatment effects. These biases relate to the number of centres in which the study is conducted, the number of participants included in the study and the type of sponsor of the trial.

With respect to the number of centres, meta-epidemiological studies have shown that single centre studies overestimate treatment effects relative to multicentre studies (ROR 0.73, 0.64-0.83), even after adjustment for differences in the reporting of random sequence generation, allocation concealment, blinding, management of loss to follow up, sample size and publication bias (which is where studies with statistically significant findings are more likely to be published than studies with statistically non-significant findings) (28). There are many possible reasons why single centre studies find larger treatment effects than multicentre studies. First, single centre studies are often conducted by highly specialized centres with a keen interest in the intervention being examined in the clinical trial (28). If the delivery of this intervention requires a highly specialized and well trained group of investigators, it may be difficult for multicentre trials to replicate the treatment effects observed in single centre trials. Furthermore, single centre trials may include a more homogenous population, whose response to the intervention may be more robust than when the intervention is applied to a more general population (28,29). Given that multiple clinical interventions that have been supported by single centre trials have

subsequently been overturned by multicentre trials (29), it is important to consider the number of centres in a clinical trial when interpreting the results.

Other meta-epidemiological studies have shown that clinical trials with small sample sizes were more likely to overestimate treatment effects compared to clinical trials with large sample sizes (ROR 0.77, 0.65 to 0.91 for the lowest quarter of sample size compared to the highest quarter of sample size) (30). The ROR remained consistent even after adjustment for methodological characteristics. The mechanism underlying this difference in treatment effect by sample size may also be due to the homogeneity of the population included in smaller studies compared to larger studies. Alternatively, it may be a consequence of publication bias, which is when studies with non-significant results are less likely to be published (30).

Finally, the funding source of clinical trials has also been shown to influence study findings. Recently, researchers conducted a systematic review and meta-analysis of observational studies that compared industry funded to non-industry funded studies. Based on 48 observational studies, industry funded trials were more likely to have favorable efficacy results (risk ratio 1.24, 1.14-1.35), harm results (risk ratio 1.87, 1.54-2.27) and conclusions (risk ratio 1.31, 1.20-1.44) (31). There was also no difference in the reported methodological quality of industry funded trials and non-industry funded trials with respect to random sequence generation, allocation concealment, blinding and loss to follow up (31). The reasons why industry funded studies may have favorable efficacy results are also multifactorial. For instance, investigators have shown that in comparative effectiveness trials, industry funded studies often deliver the investigational drug at a higher dose than the comparator drug (32). Alternatively, the comparator drug is given outside of its

recommended therapeutic range. Both of these practices may lead to favorable results for the investigational drug. Another contributing factor may be publication bias, particularly as multiple studies have shown that pharmaceutical companies suppress publications that do not favor their investigational medicines (33).

Taken together, when the results of clinical trials are evaluated, the characteristics of the study must be considered along with the methodological design and conduct of the trial.

1.5 LIMITATIONS OF METHODOLOGICAL STUDIES OF PUBLISHED TRIALS AND THE NEED FOR CLINICAL TRIAL PROTOCOLS

Despite the compelling findings of meta-epidemiological studies, there are important limitations to consider. First, meta-epidemiological studies are still observational in design and may be subject to confounding (34). Second, meta-epidemiological studies have relied on the final publication of RCTs, which may not be reflective of the actual conduct of the study. Indeed, researchers may have actually used good methodological strategies but did not describe it in the publication. The final publications may therefore inaccurately portray the actual methodological rigour of the trial.

To address this limitation and to identify strategies to improve clinical trials, researchers have sought alternative, more detailed documentation about the conduct of clinical trials. In the last decade, trial protocols have become an integral component of this effort. Clinical trial protocols are important roadmaps for the conduct of trials (35). The

methodology in the protocol must be finalized before the study is initiated and approved by a research ethics committee (REC). Although post-hoc alterations can occur, they should be clearly documented as an amendment to the study protocol.

By evaluating trial protocols, researchers can obtain a more conclusive assessment of the methodological conduct of trials. The first assessment of protocols involved an unselected cohort of 102 protocols for RCTs approved by RECs in Denmark in 1994 and 1995. The authors assessed the reporting of allocation concealment in the trial publications and compared it to the study protocol. Allocation concealment was classified as adequate, unclear and inadequate. Notably, 96 of 102 RCTs (94%) had unclear allocation concealment within the final publication (3). When compared to the protocol, 15 of these 96 RCTs had adequate allocation concealment (16%), 80 had unclear allocation concealment (83%) and one had inadequate allocation concealment (3). This suggests that there is reasonable concordance between methodological items reported in the trial publication and the trial protocol, with only a modest amount of misclassification. It also reinforces the findings of meta-epidemiological studies.

Notably, access to clinical trial protocols have also revealed an additional source of bias that was previously unrecognized. This bias, known as selective outcome reporting, is the post-hoc inclusion or exclusion of results in a clinical trial based on the statistical significance. Using the aforementioned 102 clinical trial protocols and the corresponding publications, the authors first classified each outcome in the publication based on whether it was fully reported or incompletely reported. Full reporting of outcomes was defined as the provision of sufficient information such that the outcomes could be included in meta-analysis (2). The authors also assessed whether the outcome was statistically significant in

the trial publication and conducted a meta-analysis to determine whether completeness of reporting was associated with statistical significance (2). Of interest, the authors found that efficacy and harm outcomes were significantly more likely to be completely reported if the result was statistically significant (odds ratio 2.4, 1.4-4.0 for efficacy outcomes and odds ratio 4.7, 1.8-12.0 for harm outcomes) (2).

Furthermore, among the 92 trials that defined a primary outcome in the trial protocol, the authors noted that 51 (62%) had a major discrepancy in the reporting of the outcome between the protocol and the trial publication (2). A major discrepancy was defined as a change in the priority of a primary outcome to secondary or unspecified, the upgrade of a secondary outcome to primary, omission of a primary outcome from the published report, or the addition of a new outcome to the publication that was not pre-specified in the study protocol (2). Sixteen of these 51 trials also had discrepancies that favored statistically significant outcomes (2). When the authors repeated their study in a cohort of government funded clinical trials, the results were unchanged (36). Therefore, selective reporting of outcomes was not limited to industry funded studies.

Taken together, there are a wide range of methodological and non-methodological factors that may alter the findings of clinical trials. Each of these factors are most reliably discerned from clinical trial protocols but assessments of trial publications can also provide insight into whether the treatment effects in the clinical trial are overestimated.

1.6 IMPROVING THE DESIGN OF CLINICAL TRIALS AND CONSORT

Given the wide-ranging sources of bias in clinical trials, efforts were initially directed towards improving the transparent and complete reporting of the design and conduct of clinical trial publications. The first effort in this regard was to develop a minimum set of recommendations for the reporting of randomized trials in journal articles. This initiative, known as the Consolidated Standards of Reporting Trials (CONSORT), was introduced in 1996 (37) and endorsed by multiple medical journals. The CONSORT statement has also been updated twice, most recently in 2010.

At its inception, the impact of CONSORT on the reporting of clinical trials was evaluated (38). In a before-after analysis, investigators assessed the reporting of the CONSORT checklist items in four high impact journals: the *New England Journal of Medicine*, *The Lancet*, *The Journal of the American Medical Association* and the *British Medical Journal* (38). The *New England Journal of Medicine* did not endorse CONSORT at the time of the study while the remaining journals endorsed the initiative (38). Compared to 1994 (two years before CONSORT was introduced), the frequency of reporting of random sequence generation, allocation concealment and blinding increased in all journals in 1998. However, the increase was statistically significant for CONSORT endorsing journals only (38).

A comparable study examined trends in methodological reporting of RCTs in high impact general medical and specialty journals. The investigators examined 284 RCTs published between 1995 and 2010 and found that the reporting of allocation concealment improved over time (32.3% in 1995, 31.9% in 2000, 42.1% in 2005, and 58.4% in 2010, p-value for trend=0.003) (39). The authors examined other methodological characteristics – namely random sequence generation, blinding and attrition – within the framework of a

composite quality score (39). Although composite quality score are no longer recommended, it is noteworthy that the authors also found that the proportion of RCTs with high quality scores increased over time (29% in 1995, 42% in 2000, 48.7% in 2005, 51.9% in 2010, p=0.038) (39).

CONSORT has been endorsed by other journals that may not be considered "high impact" but the improvement in reporting across the spectrum of clinical trials has been rather modest. In particular, investigators assessed trends in a representative sample of RCTs published in PubMed indexed journals, without restriction to a medical specialty, disease type, geographic location or journal. In December 2000, 519 RCTs were identified whereas in December 2006, 616 RCTs were identified (10). Despite the increase in the number of RCTs, the reported methodological conduct showed only marginal improvement between the two time points (10). In the December 2000 cohort for example, 21% of RCTs described adequate methods for random sequence generation whereas 34% described adequate details in the December 2006 cohort (10). Similar patterns were noted for methods of allocation concealment (18% vs. 25%) and blinding (29% vs. 26%) (10).

Recognizing the limitations of assessing published reports of clinical trials, these findings suggest that the methodological quality of RCTs may have improved in high impact journals but the improvement in lower impact journals – wherein the majority of RCTs are published – may be limited.

1.7 IMPROVING THE DESIGN OF CLINICAL TRIALS AND THE STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS (SPIRIT) INITIATIVE

With only modest gains from CONSORT and the recognition that poor methodological quality in a trial publication may reflect poor methodological quality in the trial protocol, efforts have now been directed towards improving clinical trials when they are first designed. This approach is also pragmatic because many of the strategies discussed to prevent bias are easy to implement and not costly. For instance, investigators analyzed a sample of 200 published trials to determine whether improvements to trial methodology at the design stage could reduce bias in these studies (40). One hundred and forty-two of these trials were deemed by systematic reviewers as being at high risk of bias and an expert panel on trial methodology identified that adjustments to reduce bias in these trials was possible for 136 (96%) of the studies (40). Easy adjustments to trial methodology that could reduce the risk of bias at no cost were possible in 71 of the 142 studies (50%) (40). Focusing on the design stage of clinical trials may therefore be a robust opportunity to improve the design of clinical trials and reduce bias.

In response to the deficiencies in the design and conduct of clinical trials, the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) Initiative was introduced (5,6). The aim of SPIRIT is to improve the completeness of clinical trial protocols by focusing on the design stage of trials and producing evidence-based recommendations for a minimum set of checklist items to be addressed in clinical trial protocols. The checklist includes methodological items discussed in this introductory chapter as well as other aspects of trial design such as planning of the intervention, recruitment, and planned dissemination strategies. The SPIRIT checklist was published in 2013 and the utility of the checklist must be evaluated.

The first step in the evaluation – and herein the basis of this Doctor of Philosophy (DPhil) thesis – is to establish a baseline measure of the quality of recent clinical trial protocols prior to the publication of the SPIRIT guideline in 2013. Through the support of the United Kingdom (UK) National Research Ethics Service (NRES), the Centre for Statistics in Medicine (CSM) has access to clinical trial protocols for research institutions in the UK. The trial protocols forming this cohort– entitled the ASPIRE (Adherence to SPIrit REcommendations) study – were approved between January and December 2012. Access to this cohort of clinical trial protocols also provides a unique opportunity to conduct longitudinal studies on the methodological conduct of contemporary clinical trials. it may also provide insight into new areas in clinical trial design and conduct that are not currently considered in SPIRIT and may warrant inclusion.

2

METHODS

This thesis will provide an overview of the ASPIRE cohort and a baseline assessment of the quality of clinical trial protocols using the SPIRIT checklist. The project will also address research questions related to the following aspects of clinical trial conduct: Selection of Outcomes; Transparency and Accountability in the Conduct of Clinical Trials; and two special topics: Design and Conduct of Multi-Arm Trials and the Design and Conduct of Phase I Trials. The specific research questions corresponding to each of the aforementioned thematic areas are detailed below.

2.1 RESEARCH QUESTIONS

Theme: Overview of the ASPIRE Cohort and Quality of Protocols

Chapter Three: What are the general characteristics of clinical trials approved by the UK NRES in 2012? What percentage reported each item on the 2013 SPIRIT checklist and what

is the adequacy and completeness of reporting on methodological items (eligibility criteria, interventions, primary outcome measures, study design and sample size)? What factors – among funding source, intervention type, sample size and number of centres – predict adequate reporting of methodological items?

Theme: Selection of Outcomes

Chapter Four: Among trial protocols approved by the UK NRES in 2012, what percentage included PROMs when necessary and what factors – among funding source, intervention type, sample size and number of centres – are associated with the use of PROMs?

Theme: Transparency and Accountability

Chapter Five: Among trial protocols approved by the UK NRES in 2012, what percentage were registered in a trial registry? What is the number and type of discrepancies in methodological items between the trial protocol and trial registry? What factors – among funding source, intervention type, sample size and number of centres – predict discrepancies? What were reasons for which trials were not registered (qualitative study)?

Theme (Special Topics): Epidemiology and Design of Multi-Arm Trials

Chapter Six: Among trial protocols for multi-arm studies approved by the UK NRES in 2012, what is the structure of multi-arm trials and what percentage considered all trial arms in the study objectives and planned analysis of the trial? What percentage reported details of adjustment for statistical comparison (global test for comparison and adjustment of the type I error rate for pairwise comparisons) and did reporting vary by funding source?

Theme (Special Topics): Design, Conduct, and Dissemination of Phase I Trials

Chapter Seven: Among trial protocols for phase I trials approved by the UK NRES in 2012, what percentage reported appropriate methods for the calculation of starting doses? What percentage reported an interval of observation for dosing of the first participant and subsequent participants? What percentage of phase I trials were registered in a trial registry compared to phase II-IV trials? What percentage of phase I trials were disseminated compared to phase II-IV trials?

2.2 RESEARCH ETHICS APPROVAL IN THE UNITED KINGDOM

Research ethics approval in the UK is a centralized process (1). There are five regional research ethics offices – Bristol, Manchester, Harrow, Nottingham and London – each of which oversee multiple RECs within their jurisdictional area (1). The Bristol office is providing access to clinical trial protocols for the ASPIRE study and services to London, Home Counties and South West (1). Centralized processing of ethics applications allows for some standardization of ethics review. For instance, all studies applying for approval must complete an Integrated Research Application System (IRAS) form. The IRAS form is an extended document where investigators provide general information about the design of their study and respond to specific questions about its methodological conduct (2). This includes important methodological items such as how the randomization was performed, identification of the primary and secondary outcomes and reporting of sample size calculations. Other elements related to clinical trial conduct must also be reported including

planned strategies for dissemination of the results. Centralized ethics review also ensures that any new policies adopted by the NRES are immediately applied to all clinical trials in the UK.

2.3 DATA SOURCES AND STUDY SELECTION CRITERIA

The trial protocols forming the ASPIRE cohort were approved between January and December 2012. Details of all clinical trials receiving ethics approval in 2012 were obtained from the Bristol office of the NRES, which is responsible for 19 RECs. Based on this information, study titles were screened in single by the DPhil candidate to exclude studies that were clearly not clinical trials. For the remaining studies, the IRAS form that must accompany each ethics application was screened. A review of the full protocol and IRAS form was conducted if the investigator identified the study under one of the following IRAS categories:

1. Clinical trial of an investigational medicinal product (CTIMP)
2. Clinical investigation or other study of a medical device (CIMD)
3. Combined clinical trial of an investigational medicinal product and an investigational medical device (CTIMP and CIMD)
4. Other clinical trial to study a novel intervention or to compare interventions in clinical practice.

For these four types of studies, the full IRAS form and trial protocol was reviewed to confirm that it was in fact a clinical trial and that the protocol version approved by the REC could be obtained. The date that ethics approval was granted was ascertained from letters of correspondence between RECs and researchers. These letters also detailed the version number and date for the approved protocol, which allowed for identification of the correct trial protocol for this thesis.

Controlled and uncontrolled (single intervention with no control group) clinical trials were included, as well as NRTs and RCTs and phase I trials. RCTs were defined as a prospective study that assessed two or more healthcare interventions that were randomly assigned to human participants. Clinical trials investigating medicinal products, medical devices, and other non-medicinal interventions were included. Studies primarily evaluating physiology or health economics were excluded.

2.4 CHALLENGES IN THE PROCESS OF ACQUIRING CLINICAL TRIAL PROTOCOLS

Over the last decade, RECs have become more willing to provide researchers with access to clinical trial protocols (3). Nonetheless, important challenges remain. In the case of the ASPIRE cohort, this study was the first time the UK NRES provided this level of access to clinical trial protocols. Therefore, extensive negotiations were required between the CSM and the NRES to determine the nature of access that would be granted and to clarify the expectations of both parties.

After approval was granted, it was quickly apparent that the NRES lacked the infrastructure to readily identify clinical trial protocols. Although ethics applications were processed through an electronic database, the research documents corresponding to each application were not consistently uploaded to the database and a large percentage of applications were missing a protocol. If protocols were uploaded, they were included in multiple files with all other components of the ethics application, such as letters of correspondence between the committee and researchers, amendments, investigator brochures, consent forms and other documentation. Therefore, identifying the clinical trial protocol that was approved by RECs was an iterative process, which required a detailed review of the complete ethics application for each potential clinical trial and multiple correspondences with the RECs to obtain the right documentation.

2.5 DATA ABSTRACTION

For the purposes of this thesis, data abstraction was performed in single. However, duplicate data abstraction is currently ongoing using pre-piloted electronic report forms. For each trial, general characteristics, including randomization, medical specialty, number of centres, sample size and funding type were extracted. The protocol was considered as the primary source of information and the IRAS form was used for data extraction if specific details were missing from the trial protocol. Detailed selection criteria, data abstraction and methods for each study are provided within each results chapters of this thesis. At the time of this thesis preparation in August 2017, duplicate data extraction was completed for the

general study characteristics. The agreement between the DPhil candidate and other data extractors ranged from 81% to 97% and the kappa statistics was consistently over 0.7.

2.6 SAMPLE SIZE CALCULATION

A formal sample size calculation was not performed for the ASPIRE study. However, the focus was on ensuring a sufficient number of protocols were obtained to allow for estimation of the quality of reporting in clinical trial protocols. Prior analyses of trial protocols have included around 100 trials and it was empirically decided that at least twice the sample size of these previous studies (n=200) would be sufficient to obtain a baseline assessment of the quality of reporting in clinical trial protocols prior to the introduction of the SPIRIT guidelines. Furthermore, the only research question that was established at the time of this decision regarding the sample size was the study related to trial registration. A sample size of at least 200 trial protocols was also considered sufficient for the trial registration study.

2.7 DATA ANALYSIS AND STATISTICAL CONSIDERATIONS

Descriptive Analyses

The first analytic method employed in this thesis was descriptive analyses. Categorical variables were summarised using percentages and where needed, 95% confidence intervals using the logit transformation to keep the intervals between the range of 0 and 1. Normally distributed continuous variables were summarised using the mean and

standard deviation whereas non-normal variables were summarized using the median and inter-quartile range. Comparisons were made using a t-test or Wilcoxon rank sum test.

Multilevel Models

Univariable and multivariable regression analyses were used to examine the association between pre-specified predictors – funding type (industry versus non-industry), type of intervention (pharmacologic versus non-pharmacologic), sample size and number of centres (single versus multicentre) – and outcomes of interest. Relative risk estimates (risk ratios) and their associated 95% CIs are presented. P-values are also presented but were interpreted in the context of the relative risk estimate and CI as opposed to using a strict alpha cut-off.

Notably, standard regression methods could not be used due to the underlying structure of the data. This is because approval of individual clinical trial protocols in the UK are managed centrally. As previously mentioned, a network of RECs were responsible for approving protocols received from their local academic institutions. This resulted in an underlying clustered structure to the data such that clinical trial protocols from a given REC were likely to be similar to one another in quality and methodological conduct. This similarity among the trial protocols in the ASPIRE cohort is frequently referred to as correlation and it has important implications for the analysis in this thesis. In particular, cohort studies with correlated data provide less information about the underlying population for which standard errors are to be estimated (4). The correlated data points are inherently similar and therefore less representative of the population of interest, compared to if the data points were random and independent of one another (4). Given that standard errors are

calculated under the assumption that data points are independent, calculations based on correlated data must be adjusted for violating this assumption. Without this adjustment, the type 1 error rate is increased.

The degree of adjustment is dependent on the extent to which the data are correlated to one another, which is quantified as the intra-class correlation coefficient (ICC). The ICC ranges from zero (which is equivalent to independent samples) to one (which is equivalent to perfect correlation). Multiple approaches are available to obtain correct standard errors.

First, the standard error from an independent samples analysis can be corrected to account for the clustering. This method involves obtaining robust standard errors through adjustments such as the Huber-White estimator. While this approach is beneficial, it may not wholly correct for clustering because the degrees of freedom used in the analysis correspond to the number of individual observations as opposed to the number of clusters (4).

The second method is that the cluster variable can be included in the regression model as a random intercept. This is done through multi-level hierarchical models. Unlike the Huber-White standard errors, this method accounts for clustering within the regression model (4). For the purposes of this thesis, multi-level hierarchical models were used to perform univariable and multivariable regression analyses. Specifically, a random intercept model was used because it was anticipated that the quality of the clinical trial protocol and the conduct of the trial would vary between RECs. It was also anticipated that the effects of predictor variables would remain consistent across the RECs.

Measures of Relative Risk

The outcomes examined in regression analyses in this thesis will be binary. This means that for each trial, the outcome will be coded as "0" if the event of interest did not occur and "1" if the event of interest occurred. The risk of the event in one group compared to another group can then be summarized as the risk ratio. Using a two by two table, the risk ratio is derived as follows (5):

TABLE 2.1: Calculation of Risk Ratios

	Group A	Group B
Event	a	b
No Event	c	d

Risk ratio of event in Group A versus Group B = (a / a+c) / (b / b+d)

If there is no difference in the average risk between the groups, this calculation would result in a risk ratio of 1. An increase in the average risk of the event in Group A would result in a value greater than 1 whereas a decreased risk would result in a value less than 1. Given that the risk ratio directly compares the average risk of an event in one group to the other group, it is often referred to as a relative risk. For the purposes of this thesis, the risk ratio will be used as opposed to the odds ratios. The odds ratio is often used because it can be readily derived through a generalized linear model known as logistic regression. The odds-ratio is defined as follows based on table 2.1 (5):

$$\text{Odds ratio} = \text{odds in exposed} / \text{odds in unexposed}$$

$$= (a / c) / (b / d)$$

OR it can be further expressed as

$$\text{Odds ratio X odds in unexposed} = \text{odds in exposed}$$

The logarithm of both sides of this equation can be taken and subsequently expressed as a linear model with a logarithmic transformation.

$$\log(\text{odds in exposed}) = \log(\text{odds ratio}) + \log(\text{odds in unexposed})$$

The log (odds in exposed) and log (odds in unexposed) are constants. In contrast, the log (odds ratio) expresses the change in the odds of disease in the exposed versus the unexposed group. This is the basis for logistic regression.

Odds ratios are intended to approximate the risk ratio (6). However, odd-ratios overestimate the risk ratio when the incidence of an event is common (≥10%) and when the difference in the risk between the two groups is large (odds-ratio ≥2.5 or odds-ratio ≤0.5) (6,7).

Unlike the odds-ratio, there is no dedicated generalized linear model specific for the risk ratio. Therefore, one method to obtain a risk ratio is to conduct a logistic regression and convert the “over-estimated” odds-ratio to a risk ratio using a mathematical correction (6). However, this strategy may yield biased estimates and in the setting of multilevel modelling, it would not be feasible.

An alternative strategy to obtain risk ratios is to use poisson regression (8,9). Poisson regression is a form of generalized linear modeling used when the outcome is count data. This would allow estimation of the incidence rate ratio for the outcome of interest based on the predictors included in the model. Poisson regression can be extended to binary data and because the data is made up of "0"s and "1"s, the estimate from the poisson regression would approximate the risk ratio. In simulation studies, the variance estimate obtained through poisson regression is too large because the error term is for count data (8). However, this can be readily corrected by an adjustment in the variance estimator for the poisson model known as the sandwich error estimator (8). This modified poisson regression was used in this thesis to obtain relative risks.

2.8 INCLUSION OF OTHER STUDY COHORTS

Swiss Cohort: Collaboration with Researchers Developing a Swiss Cohort of Clinical Trials Protocols

Although the ASPIRE study is conducted as an independent stand-alone research project, a collaboration was established with researchers in Switzerland who examined a cohort of clinical trial protocols approved by RECs in their jurisdiction in 2012. This collaboration was particularly important for increasing the sample size in Chapter Six (the analysis of the design, conduct and analysis of multi-arm trials). In this chapter, the number of multi-arm trials available in the ASPIRE cohort (n=21) were not sufficient for robust statistical inferences. A brief review of the research ethics approval process in Switzerland is provided. In 2012, there were 9 RECs in Switzerland as well as a national research ethics

council. All studies seeking research ethics approval in Switzerland must obtain approval from the national research ethics council as well as the RECs overseeing the jurisdiction where the largest number of participants would be recruited. The Swiss cohort was made up of clinical trial protocols from 7 of the 9 RECs in Switzerland. The 2 RECs that did not contribute clinical trial protocols processed the smallest number of clinical trials of the 9 RECs. At the time of preparation of this thesis, data on 32 trials from 4 Swiss RECs were available for inclusion, with data from the remaining RECs to be forthcoming as part of the ongoing collaboration. Therefore, the data that will ultimately be obtained from the Swiss cohort can be considered representative of RCTs conducted in Switzerland.

Epidemiological Study of Randomized Trials (ESORT)

ESORT was a methodological study of the design, conduct and analysis of RCTs that were published and indexed in PubMed in December 2012. The study was designed and coordinated by the DPhil candidate and was intended to serve as an additional data source during instances where insufficient data was obtained from the ASPIRE cohort. ESORT was particularly important for increasing the sample size in Chapter Six (the analysis of the design, conduct and analysis of multi-arm trials). The RCTs in ESORT were identified through a search of PubMed using a modified version of the Cochrane highly sensitive search strategy (phase 1 search strategy with added terms: “cross-over studies” and “multicentre study”). This identified 4190 primary reports of RCTs published in December 2012 and indexed in PubMed by 17 November 2013. The full search strategy is given below.

1. (randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double blind method [mh] OR single blind method [mh] OR cross-over studies [mh] OR multicenter study [pt]) NOT (animal [mh] NOT human [mh])
2. Limit to 2012-12-01 [pdat] to 2012-12-31 [pdat]

RCTs were defined as prospective studies that assessed healthcare interventions in human participants, or groups of participants, who were randomly allocated to study interventions. Cost effectiveness analyses, interim analyses, diagnostic studies, pharmacokinetic studies, pharmacodynamic studies, and physiology studies were excluded. 1122 RCTs were identified as part of ESORT.

The DPhil candidate reviewed all titles and abstracts of retrieved records, identified studies meeting the inclusion criteria for RCTs and performed all data extraction in single. Duplicate screening and data extraction was additionally performed by a team of 8 reviewers.

3

RESULTS

THEME: OVERVIEW OF THE ASPIRE COHORT AND QUALITY OF REPORTING OF PROTOCOLS

General Characteristics of Trials in the ASPIRE Cohort and Prevalence of Reporting of SPIRIT Items in Trial Protocols

Clinical trials are considered the gold standard of research studies (1). They allow researchers to assess causal relationships between interventions and outcomes and allow for robust estimates of treatment effects. However, clinical trials are not a panacea and careful planning is required to conduct the trial efficiently and to obtain valid and generalizable results.

Although the validity of clinical trial results and the potential for bias has been the primary focus of methodological research – and indeed the primary focus of the introductory chapter to this thesis – other elements of clinical trial conduct should be

considered at the design stage of the trial. These include how the intervention and comparator will be administered in the trial, clarifying the relationship between funders and sponsors with respect to conduct and dissemination of the trial, and identifying a target population that can be successfully recruited for the trial (2-6).

Improving methodological issues in clinical trials is most readily achieved at the design stage. Although, studies of clinical trial protocols have revealed remarkable variability in the reporting of methodological items (3,7-13), most assessments have relied on protocols that were over a decade old or have been limited to protocols from a single medical specialty or funder. Furthermore, the assessments have often been focused on a few elements of clinical trial design, such as random sequence generation, allocation concealment and blinding, without consideration of other relevant topics such as recruitment. It is therefore unclear whether the results from previous assessments of clinical trial protocols are generalizable to a contemporary cohort of trials, irrespective of medical specialty and funding source.

The primary objectives of this thesis chapter were four-fold. First, it was to compare the characteristics of studies approved by the Bristol Regional Office to studies approved by other regional offices. Second, it was to provide an overview of the clinical trials included in the ASPIRE cohort. Third, it was to assess the quality of reporting in clinical trial protocols using the items specified in the SPIRIT checklist. Fourth, it was to determine what study characteristics were associated with more frequent reporting of important methodological items (specified below).

3.1 CHAPTER SPECIFIC METHODS

Assessment of General Characteristics of All Studies Approved in the Bristol Regional Office and Other Regional Office United Kingdom in 2012

In order to compare the volume and characteristics of studies approved by the Bristol Regional office to other regional offices in the UK, a list of study titles and general application details were assessed for research applications receiving ethics approval in 2012. The focus was on study titles because the HRA only approved the ASPIRE cohort to include clinical trials from the Bristol Regional office. Therefore, detailed ethics applications for other regional offices were not available. The specific items ascertained from the study title and general application details were the REC responsible for approval; the intervention type, defined as a clinical trial of an investigational medicinal product (CTIMP) or a non-CTIMP; the study phase, defined as phase 1 or phase II-IV; the clinical area, defined as pediatric and non-pediatric and the funding source, defined as solely or partially industry funded and non-industry funded.

Assessment of General Characteristics of the ASPIRE Cohort

Studies approved by the Bristol Regional office of the HRA were assessed for inclusion in the ASPIRE cohort. The subset meeting the inclusion criteria for the ASPIRE cohort were assessed for additional general characteristics. The specific items that were extracted were the clinical area, defined as medical, paediatric or surgical; whether the trial

was a pilot study; the design of the trial, defined as parallel group, crossover, factorial, split body or other; the intervention employed in the trial, defined as drug, device, procedure, behavioural intervention or other; the number of centres, defined as single, multiple or unclear; the number of study arms included in the trial; the sample size of the trial and the funding source for the trial, defined as solely or partially industry funded and non-industry funded. The classification of these items, and particularly the medical area, was based on agreement with collaborators to allow for uniform data extraction.

Inclusion Criteria

The assessment of the quality of reporting in clinical trial protocols was performed for RCTs. NRTs and pilot studies were excluded. These exclusion criteria were applied for two reasons. First, although SPIRIT may be used in the preparation of protocols for all clinical trials, there are many elements in the checklist that may not apply to NRTs and pilot studies but there is no consensus on these items. Including pilot studies and NRTs would potentially falsely underestimate the quality of reporting in clinical trial protocols. Second, these exclusion criteria were applied because of the collaborative nature of the ASPIRE study. In particular, the evaluation of the SPIRIT checklist is part of a multinational initiative and collaborators from other countries – namely Switzerland, Germany, and Canada – were only able to gain access to protocols for RCTs. The identification of which SPIRIT items to assess and how to operationalize the extraction process was done in collaboration with these investigators, and it was therefore considered pragmatic to focus on the type of clinical trial protocols that were accessible across all sites.

Data Extraction

The complete SPIRIT checklist includes 270 individual items grouped under 33 separate headings. For instance, the heading "sample size calculation" has nine items which are relevant to the calculation of sample size, including but not limited to the statistical test used to calculate the sample size, type I error, type II error, and the minimum anticipated difference or event rate. Given the size of the checklist, it was not pragmatic to include all items. The process of identifying which SPIRIT items to include in data extractions was as follows. First, all items were included where the heading related to the formulation of a research question using the PICO acronym (Population, Intervention, Comparator and Outcome). These were headings that defined the target population, defined the intervention used and any comparators and defined the outcomes of the study. These headings were considered important because they are relevant to defining the research question in the clinical trial. Likewise, all SPRIT items were included for headings related to sample size calculation, random sequence generation, allocation concealment, and blinding. These headings were also chosen because of their importance for reducing bias in clinical trials.

For the remaining SPIRIT headings, two investigators – one of whom was the DPhil candidate – independently, reviewed each of the items under the heading and selected one or two items that encompassed the core message of the heading. These selections were compared and reviewed with two additional collaborators from other sites to achieve consensus on the final selection. This produced a compressed 83 item SPIRIT checklist with at least one item for each SPIRIT heading. This checklist was piloted on five trial protocols to develop operational definitions for each of the items (Appendix 1).

Each trial protocol was assessed to determine whether each of the 83 items was addressed (Yes or No). "Not Applicable" was also used for some items that were not relevant for a specific clinical trial. For instance, the SPIRIT item related to the name and dosage of a pharmacological agent would not be relevant to a clinical trial protocol involving a non-pharmacological intervention.

Statistical Analysis

A descriptive analysis was performed, as detailed in the Methods Chapter. To aid in the interpretation of the results, the 83 SPIRIT items were assessed as follows: The items that were reported in at least 95% of trial protocols and in less than 5% of trial protocols were identified. This was limited to items that were applicable to at least 25 trials. For the SPIRIT headings where all items were assessed (e.g. heading related to the PICO for the research question), the results were further summarized by grouping the items based on whether they were frequently reported (defined as reported in ≥80% of protocols) or infrequently reported (<80% of protocols).

Multilevel poisson regression was used to assess the association between general study characteristics and the reporting for specific SPIRIT headings. Reporting and non-reporting was assessed at the level of each SPIRIT heading. Therefore, when all SPIRIT items for a particular heading were reported, the protocol was coded as reported. If at least one item was not reported, the protocol was coded as not reported.

The SPIRIT headings examined as outcomes in multivariable models were limited to SPIRIT headings where all items in the checklist were assessed. With this criteria, the SPIRIT headings examined were those related to the eligibility criteria, intervention and

comparator, the study outcome, blinding, sample size calculation, random sequence generation and allocation concealment.

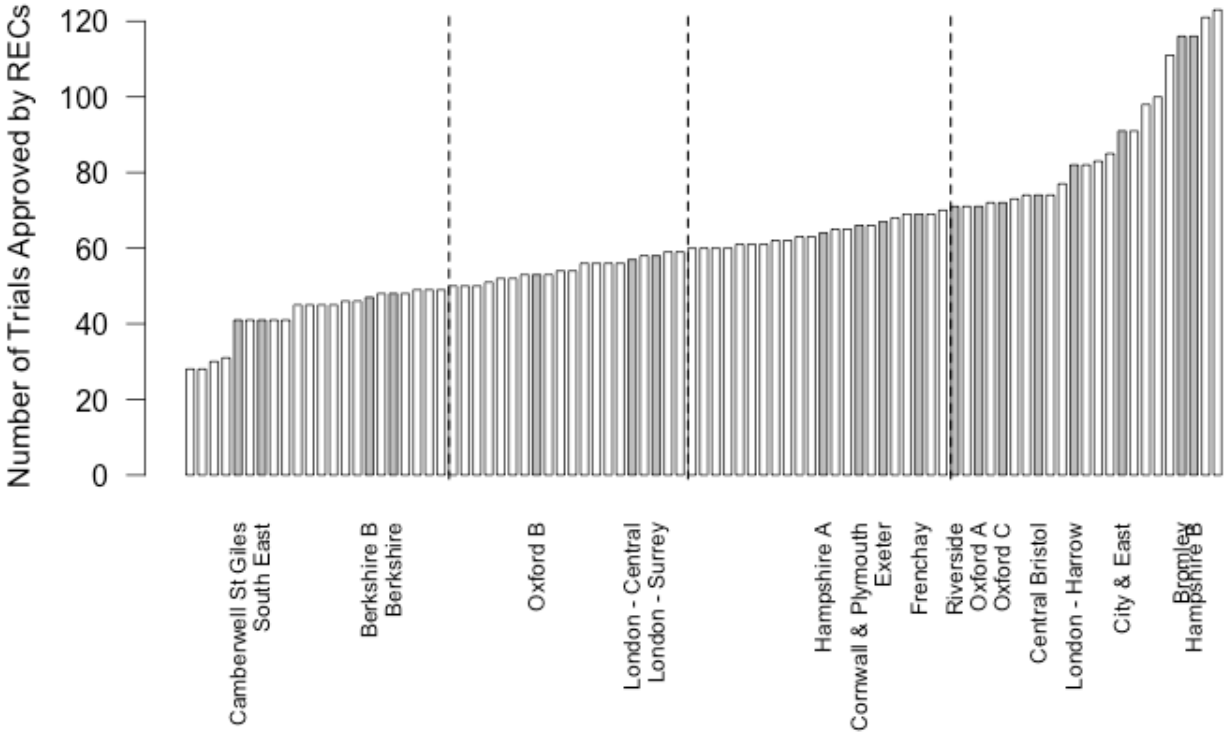
The model accounted for clustering at the level of the REC. The predictors of interest included in multivariable models were funding source (industry versus non-industry), intervention type (pharmacologic versus non-pharmacologic), the number of centres (multiple versus single centre) and sample size (continuous).

3.2 CHAPTER RESULTS

Among 5459 trials approved by RECs in the UK in 2012, 1247 studies (23%) received ethics approval from the Bristol Regional office of the NRES and 4212 received ethics approval from a non-Bristol Regional office. With respect to the volume of applications approved by individual RECs, eight of the 19 RECs in the Bristol office were in the top quarter (Figure 3.1). The Bristol Regional office was similar to non-Bristol offices with respect to the number of CTIMPs, the phase of the studies conducted, the clinical area and the funding source of the studies (Table 3.1).

Of these 1247 studies, 873 were excluded because they were not clinical trials (Figure 3.2). IRAS forms and trial protocols (when available) were reviewed for 374 studies and a further 86 studies were excluded as they were not clinical trials.

Figure 3.1: Distribution of Studies Approved by Research Ethics Committees in the United Kingdom in 2012



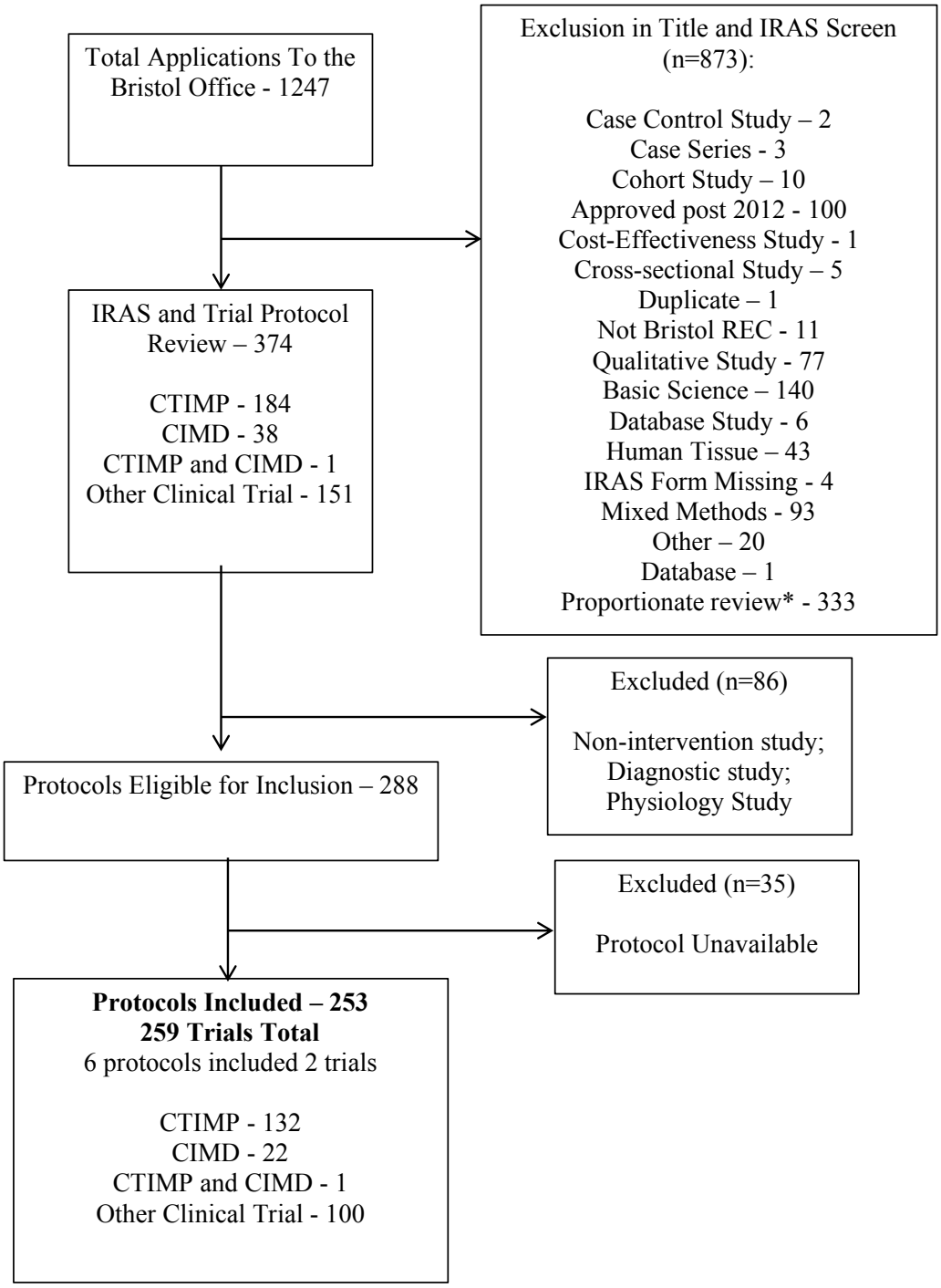
Names of the 19 research ethics committees belonging to the Bristol Office are indicated on the graph and their corresponding bars are colored grey. Vertical dashed lines represent the 25th, 50th, and 75th percentile for number of trials approved by each research ethics committee.

Table 3.1: General Characteristics of Clinical Studies at Regional Offices in the United Kingdom Based on an Assessment of Study Titles and General Application Details

Characteristic	Bristol Office (n=1247)	Non-Bristol Regional Offices (n=4213)
Intervention (%)		
CTIMP	184 (15)	505 (12)
Non-CTIMP	1063 (85)	3708 (88)
Study Phase (%)		
Phase I Studies	28 (2)	125 (3)
Non-Phase I Studies	1219 (98)	4088 (97)
Clinical Area (%)		
Non-Pediatric	1086 (87)	3775 (90)
Pediatric	161 (13)	427 (10)
Funding (%)		
Solely or Part Industry	11 (1)	35 (1)
Non-Industry	76 (6)	241 (6)
Unclear from Title Screen	1160 (93)	3926 (93)

Note that this table provides a breakdown of the characteristics of all ethics applications submitted to Bristol and non-Bristol regional offices. This includes clinical trials and non-clinical trials.

Figure 3.2: Identification of Included Studies for the Overall ASPIRE Cohort



*Studies receiving proportionate review are non-intervention studies that do not involve any material ethical issues and can be reviewed within a shortened time frame

Accordingly, the study cohort was composed of 288 ethics applications for clinical trials. Study protocols were available for 253 of the 288 identified applications and single data extraction was performed on all of these protocols. Six protocols contained two clinical trials, therefore increasing the number of studies to 259 trials.

General Characteristics of the Study Cohort

One hundred and ninety-three clinical trials (75%) were RCTs whereas 66 trials (25%) were NRTs (Table 3.2). The majority of studies in the cohort examined pharmacologic interventions (151/259, 58%), whereas 29 trials (11%) examined a device, 24 trials (9%) examined surgeries or procedures and 44 trials (17%) examined behavioural interventions. Approximately half of the trials were solely or partly industry funded (129/259, 50%). The results for the NRT and RCT subgroups were similar to the overall cohort. With respect to sample size, the median sample size was 72 (inter-quartile range [IQR] 40-200) for the overall cohort, 100 (IQR 50-274) for the RCTs and 40 (IQR 20-60) for the NRTs.

There were 35 applications with a missing trial protocol (Table 3.3). The applications were widely distributed among RECs and there were 12 RECs with at least one missing trial protocol. Studies with missing protocols were more likely to be randomized, involve a pharmacological intervention and be single centre trials. They were also larger in sample size and more likely to be industry funded.

Table 3.2: General Characteristics of Trials Included in the ASPIRE Cohort

Characteristics	All (n=259)	Randomized (n=193)	Non-Randomized (n=66)
Clinical Area (%)			
Medical	204 (79)	150 (77)	54 (82)
Paediatrics	15 (5)	9 (5)	5 (8)
Surgical	41 (16)	34 (18)	7 (11)
Pilot Study (%)	72 (28)	52 (27)	20 (30)
Design (Randomized Only, n=193, %)			
Parallel Group	167 (87)	167 (87)	-
Crossover	17 (9)	17 (9)	-
Other	9 (4)	9 (4)	-
Intervention (%)			
Drug	151 (58)	112 (58)	39 (59)
Device	29 (11)	18 (9)	11 (17)
Procedure/Surgery	24 (9)	20 (10)	4 (6)
Behavioural/Lifestyle/Education	44 (17)	33 (17)	11 (17)
Other	11 (4)	10 (5)	1 (2)
Number of Centres (%)			
Single	105 (41)	76 (39)	29 (44)
Multiple	152 (59)	116 (60)	36 (55)
Unclear	2 (1)	1 (1)	1 (2)
Number of Arms (%)			
1	66 (26)	-	61 (92)
2	148 (57)	148 (77)	4 (6)
3	25 (10)	25 (13)	-
4	12 (5)	12 (8)	-
>4	8 (3)	8 (4)	1 (2)
Sample Size (median and inter-quartile range)	72 (40-200)	100 (50-274)	40 (20-60)
Funding (%)			
Solely or Part Industry	129 (50)	94 (49)	35 (53)
Non-Industry	130 (51)	99 (51)	31 (47)

Table 3.3: General Characteristics of Trials With and Without Protocols Available

Characteristics	Applications Without Protocol Available (n=35)	Applications with Protocol Available (n=259)	Applications for RCTs with Protocol Available (n=193)
Randomized (%)	30 (86)	193 (75)	193 (100)
Intervention (%)			
Drug	26 (74)	151 (58)	112 (58)
Device	3 (9)	29 (11)	18 (9)
Procedure/Surgery	1 (3)	24 (9)	20 (10)
Behavioural/Lifestyle/Education	5 (14)	44 (17)	33 (17)
Other	0 (0)	11 (4)	10 (5)
Number of Centres (%)			
Single	26 (74)	105 (40)	29 (44)
Multiple	9 (26)	152 (59)	36 (55)
Unclear	0 (0)	2 (1)	1 (2)
Sample Size (median and inter-quartile range)	367 (56-603)	72 (40-200)	40 (20-60)
Funding (%)			
Solely or Part Industry	28 (80)	129 (50)	35 (53)
Non-Industry	7 (20)	130 (51)	31 (47)

Quality of Reporting in Clinical Trial Protocols

One hundred and twelve RCTs were included in the analysis of the quality of reporting in clinical trial protocols. Six SPIRIT items were reported at least 95% of the time (Table 3.4) and these were the version number for the trial protocol (95%, 95% CI 88%-98%); the description of the comparator in the background section of the trial protocol (95%, 95% CI 90%-98%); the objectives for each comparison (95%, 95% CI 88%-98%); inclusion and exclusion criteria for trial participants (96%, 95% CI 91%-99%); the individual administering the intervention for behavioural intervention trials (96%, 95% CI 83%-99%); and the generic name, dose and schedule for trials of pharmacological interventions (98%, 95% CI 90%-100%). Four SPIRIT items were reported in 5% or fewer of protocols: roles and responsibilities of the funder with respect to the conduct of the trial (5%, 95% CI 2%-12%); declaration of financial and other competing interests (5%, 95% CI 2%-10%); financial and non-financial incentives to encourage recruitment by investigators (0%); and planned dissemination of the full trial protocol (1%, 95% CI 0%-6%).

Table 3.4: Reporting of SPIRIT Checklist Items in Clinical Trial Protocols

SPIRIT Heading	SPIRIT Item	Number Reporting Variable of Interest (%) [95% CI] (n=112)
Title	Basic study design	38 (34) [26-43]
Trial Registration – Registry	Trial identifier	56 (50) [41-59]
Protocol	Version Number	106 (95) [88-98]
Funding	Sources of financial and non-financial support	81 (72) [63-80]
Roles and Responsibilities - Contributors	Names of protocol contributors	14 (13) [8-20]
Roles and Responsibilities – Sponsor Contact Information	Name of sponsor contact	24 (21) [15-30]
Roles and Responsibilities – Sponsor and Funder	Conduct	6 (5) [2-12]
Roles and Responsibilities – Steering Committee	General membership (n=48)	23 (48) [34-62]
Background and rationale	Research question, up-to-date systematic review	17 (15) [10-23]
Background and rationale – choice of comparators	Comparator description	107 (95) [90-98]
Objectives	Objectives for each comparison	106 (95) [88-98]
Trial Design	Type of trial	96 (86) [78-91]
Study Setting	Countries where data will be collected	64 (57) [48-66]
Eligibility criteria	Inclusion and exclusion criteria for trial participants	108 (96) [91-99]
	Inclusion and exclusion criteria for study centres (n=84)	9 (11) [6-20]
Intervention (drug)	Generic name, Dose and Schedule (n=66)	65 (98) [90-100]
Intervention (non-drug)	Setting (n=46)	43 (93) [81-98]
	Individuals administering interventions (e.g., expertise) (n=46)	44 (96) [83-99]
Interventions –	Standard criteria for modifications	48 (55)

Modifications	(n=88)	[44-65]
Interventions – Adherence	Strategies to improve adherence or any procedures for monitoring adherence (n=74)	31 (42) [31-54]
Interventions – Concomitant care	Permitted care and interventions	75 (67) [58-75]
Outcomes – Primary Outcome only	Specific measurement variable (n=107)	103 (96) [90-99]
	Analysis metric (n=85)	68 (80) [70-87]
	Time point (n=103)	88 (85) [77-91]
Participant timeline	Timing of visit	104 (93) [86-96]
Sample size	Outcome used (n=109)	92 (84) [76-90]
	Assumed values for outcome in each study group (n=109)	41 (38) [29-47]
	Rationale or reference for assumed outcome values (n=109)	47 (43) [34-53]
	Statistical test (n=109)	42 (39) [30-48]
	Alpha (n=109)	91 (83) [75-89]
	Power (n=109)	97 (89) [81-94]
	Adjustment for missing data, if relevant (n=109)	57 (51) [42-60]
	Rationale for intended sample size if not derived statistically (n=3)	3 (100) [NA]
Recruitment	Location of participant recruitment	51 (46) [36-55]
	Who will recruit participants	37 (33) [25-42]
	Expected recruitment rates	20 (18) [12-26]
	Plans to monitor recruitment during trial	15 (13) [8-21]
	Financial and non-financial incentives for participants	9 (8) [4-15]
	Financial and non-financial incentives for investigators	0 (0) [NA]
Allocation – Sequence Generation	Random allocation - Method of sequence generation	80 (71) [62-79]
	Random allocation - Allocation	81 (72)

	ratio	[63-80]
	Random allocation - Type of randomization	59 (53) [43-62]
	Non-random allocation - Method	N/A
	Non-random allocation - Rationale	N/A
Allocation – concealment mechanism	Allocation concealment mechanism	84 (75) [66-82]
Allocation - Implementation	Who will enrol/assign participants	38 (33) [26-43]
Blinding/Masking	Blinding status of participants (n=110)	97 (90) [81-94]
	Blinding status of care providers (n=110)	94 (85) [98-92]
	Blinding status of outcome assessors (n=110)	73 (65) [56-74]
Blinding – Emergency Unblinding	Circumstances/conditions when unblinding is permissible (n=110)	37 (57) [44-69]
Data Collection Methods	Data collection personnel	57 (51) [42-60]
Data Collection Methods - Retention	Strategies to promote participant retention and complete follow-up (n=104)	26 (25) [18-34]
Data Management	Data entry and coding processes	16 (14) [9-22]
Statistical Methods - Outcomes	Main analysis for primary outcome, including analysis methods for statistical comparisons	89 (82) [73-88]
	How missing data will be handled	44 (40) [31-50]
	Effect measure	-
	Significance level	48 (44) [35-54]
	Intended use of confidence intervals	-
Statistical Methods – Additional Analyses	Definition of subgroup categories (n=25)	16 (64) [43-81]
Statistical Methods – Analysis population and missing data	Which participants will be included in the main analyses	79 (71) [61-78]
Data Monitoring Committee	DMC included? If no, reason given?	59 (53) [43-62]
Data Monitoring – Interim Analyses	Who has ultimate authority to stop the trial	-
Harms	Anticipated/Unanticipated	90 (80)

		[72-87]
Auditing	Procedures and anticipated frequency of audits	14 (13) [7-20]
Protocol Amendments	Process for making amendments	35 (31) [23-41]
Consent or Assent	Consent process	100 (89) [82-94]
Consent or Assent – Ancillary Studies	Process to obtain additional consent for collection and use of data and biological specimens (n=20)	15 (75) [50-90]
Confidentiality	How data will be collected, kept secure, and maintained during the trial	87 (78) [69-85]
Declaration of Interests	Financial and other competing interests	5 (5) [2-10]
Access to Data	Who will have access to full dataset after trial completion	23 (21) [14-29]
Ancillary and post-trial care	Any plans to provide or pay for ancillary care during the trial	-
Dissemination Policy - Results	Plans to disseminate trial results to key stakeholders/publication	78 (70) [60-78]
Dissemination Policy - Authorship	Authorship eligibility criteria	24 (21) [15-30]
Dissemination Policy – Reproducible Research	Full trial protocol	1 (1) [0-6]
Informed Consent Materials	Model consent and/or assent forms	N/A (Consent forms are mandatory as per ethics application)
Biological Specimens	Details of specimen collection (n=17)	17 (100) [NA]

As for items related to the eligibility criteria, the inclusion and exclusion criteria for study centres was infrequently reported in trial protocols (Table 3.5). Likewise, half of the SPIRIT items that were relevant to the sample size calculation were infrequently reported, including assumed values for outcomes in each study group (38%, 95% CI 29%-47%), the rationale or reference for assumed outcome values (43%, 95% CI 34%-53%), the statistical test used in the sample size calculation (39%, 95% CI 30%-48%) and adjustment for missing data (51%, 95% CI 42%-60%; Table 3.4). None of the items related to random sequence generation and allocation concealment were reported in at least 80% of clinical trial protocols. Finally, the blinding status of outcome assessors (65%, 95% CI 56%-74%) and the circumstances/conditions when unblinding is permissible (57%, 95% CI 44%-69%) were infrequently reported (Table 3.5).

No association was observed between general study characteristics and the reporting of items for SPIRIT headings related to the intervention and comparator, the primary outcome and blinding (Table 3.6). However, confidence intervals were wide.

Table 3.5: Reporting of SPIRIT Checklist Items Stratified as Frequent Versus Infrequent

SPIRIT Heading	Frequently Reported Items (>80%)	Infrequently Reported Items (%) [95% CI]
Eligibility Criteria	Inclusion and exclusion criteria for trial participants	Inclusion and exclusion criteria for study centres (11%) [6%-20%]
Intervention	Generic name, Dose and Schedule Setting (non-drug) Individuals administering interventions (e.g., expertise; non-drug)	Standard criteria for modifications (55%) [44%-65%] Strategies to improve adherence or any procedures for monitoring adherence (42%) [31%-54%]

		Permitted care and interventions (67%) [58%-75%]
Outcome	Specific measurement variable Analysis metric Time point	No Items
Sample Size	Outcome used Alpha Power Rationale for intended sample size if not derived statistically	Assumed values for outcomes in each study group (38%) [29%-47%] Rationale or reference for assumed outcome values (43%) [34%-53%] Statistical test (39%) [30%-48%] Adjustment for missing data (51%) [42%-60%]
Random Sequence Generation	No Items	Method of sequence generation (71%) [62%-79%] Allocation ratio (72%) [63%-80%] Type of randomization (53%) [43%-62%]
Allocation Concealment	No Items	Allocation concealment mechanism (75%) [66%-82%] Who will enrol/assign participants (33%) [26%-43%]
Blinding	Blinding status of participants Blinding status of care providers	Blinding status of outcome assessors (65%) [56%-74%] Circumstances/conditions when unblinding is permissible (57%) [44%-69%]

Table 3.6: Association Between Study Characteristics and Reporting of SPIRIT Items

	Number of Studies	Number Reporting All SPIRIT Items (%)	Unadjusted RR (95% CI)	P-value	Adjusted RR (95% CI)	P-value
PRIMARY OUTCOME						
Intervention (%)						
Non-pharmacologic	32	19 (59)	ref	ref	ref	ref
Pharmacologic or Vaccine	75	68 (91)	1.53 (1.07-2.18)	0.020	1.33 (0.87-2.03)	0.19
Number of Centres						
Single	26	17 (65)	ref	ref	ref	ref
Multiple	81	70 (86)	1.32 (0.97-1.81)	0.082	1.21 (0.88-1.66)	0.24
Sample Size (per 1000 participants)	-	-	1.03 (1.02-1.04)	<0.001	1.03 (1.00-1.06)	0.028
Funding						
Non-Industry	52	34 (65)	ref	ref	ref	ref
Solely or Part Industry	55	53 (97)	1.47 (1.20-1.82)	<0.001	1.21 (0.96-1.53)	0.10
ELIGIBILITY CRITERIA						
Intervention (%)						
Non-pharmacologic	25	2 (8)	ref	ref	ref	ref
Pharmacologic or Vaccine	59	7 (12)	1.48 (0.31-7.03)	0.62	3.95 (0.79-19.76)	0.09
Number of Centres						
Single	4	1 (25)	ref	ref	ref	ref
Multiple	80	8 (10)	0.40 (0.05-2.98)	0.37	0.55 (0.06-4.72)	0.58
Sample Size (per 1000 participants)	-	-	0.86 (0.52-1.21)	0.44	0.92 (0.62-1.21)	0.57
Funding						
Non-Industry	37	6 (16)	ref	ref	ref	ref
Solely or Part Industry	47	3 (6)	0.39 (0.08-1.94)	0.25	0.21 (0.03-1.24)	0.08
DESCRIPTION OF INTERVENTION AND COMPARATOR						

Intervention (%)						
Non-pharmacologic	34	7 (21)	ref	ref	ref	ref
Pharmacologic or Vaccine	78	34 (44)	2.12 (0.93-4.80)	0.07	2.05 (0.86-4.86)	0.10
Number of Centres						
Single	29	11 (38)	ref	ref	ref	ref
Multiple	83	30 (36)	0.95 (0.53-1.71)	0.87	0.89 (0.49-1.64)	0.72
Sample Size (per 1000 participants)	-	-	0.96 (0.78-1.15)	0.69	0.98 (0.78-1.19)	0.87
Funding						
Non-Industry	55	16 (29)	ref	ref	ref	ref
Solely or Part Industry	57	25 (44)	1.51 (0.87-2.60)	0.14	1.05 (0.60-1.84)	0.85
BLINDING						
Intervention (%)						
Non-pharmacologic	34	18 (53)	ref	ref	ref	ref
Pharmacologic or Vaccine	78	45 (58)	1.09 (0.71-1.67)	0.69	1.04 (0.66-1.63)	0.87
Number of Centres						
Single	29	14 (48)	ref	ref	ref	ref
Multiple	83	49 (59)	1.22 (0.82-1.83)	0.33	1.31 (0.82-2.07)	0.26
Sample Size (per 1000 participants)	-	-	0.91 (0.82-0.99)	0.03	0.89 (0.81-0.97)	0.013
Funding						
Non-Industry	55	30 (55)	ref	ref	ref	ref
Solely or Part Industry	57	33 (58)	1.06 (0.70-1.60)	0.78	0.99 (0.61-1.60)	0.97
SAMPLE SIZE						
Intervention (%)						
Non-pharmacologic	34	3 (9)	ref	ref	ref	ref
Pharmacologic or Vaccine	75	14 (19)	2.14 (0.44-10.35)	0.34	2.14 (0.42-10.86)	0.34
Number of Centres						
Single	29	3 (10)	ref	ref	ref	ref
Multiple	80	14 (18)	1.70 (0.38-7.66)	0.49	1.78 (0.46-6.90)	0.40
Sample Size (per 1000 participants)	-	-	0.89 (0.72-1.05)	0.18	0.86 (0.69-1.03)	0.10

Funding						
Non-Industry	55	7 (13)	ref	ref	ref	ref
Solely or Part Industry	54	10 (19)	1.46 (0.51-4.17)	0.49	0.91 (0.36-2.31)	0.84
RANDOM SEQUENCE GENEREATION						
Intervention (%)						
Non-pharmacologic	19	1 (5)	ref	ref	ref	ref
Pharmacologic or Vaccine	45	22 (49)	9.29 (1.04-83.27)	0.046	5.07 (0.48-54.02)	0.18
Number of Centres						
Single	20	4 (20)	ref	ref	ref	ref
Multiple	44	19 (1843	2.16 (0.83-5.61)	0.11	1.69 (0.64-4.44)	0.29
Sample Size (per 1000 participants)	-	-	1.07 (0.98-1.16)	0.14	1.05 (0.99-1.10)	0.05
Funding						
Non-Industry	31	4 (13)	ref	ref	ref	ref
Solely or Part Industry	33	19 (58)	4.46 (1.53-12.99)	0.006	2.04 (0.75-5.56)	0.16
ALLOCATION CONCEALMENT						
Intervention (%)						
Non-pharmacologic	34	15 (44)	ref	ref	ref	ref
Pharmacologic or Vaccine	78	20 (26)	0.58 (0.37-0.92)	0.021	0.44 (0.17-1.13)	0.09
Number of Centres						
Single	29	5 (17)	ref	ref	ref	ref
Multiple	80	30 (36)	2.22 (0.91-5.41)	0.08	2.37 (1.06-5.28)	0.035
Sample Size (per 1000 participants)	-	-	0.99 (0.83-1.15)	0.89	0.94 (0.77-1.11)	0.10
Funding						
Non-Industry	55	19 (35)	ref	ref	ref	ref
Solely or Part Industry	54	16 (28)	0.82 (0.46-1.46)	0.50	1.29 (0.46-3.66)	0.63

3.3 CHAPTER DISCUSSION

This analysis provides an overview of the ASPIRE cohort and the quality of clinical trial protocols approved by 19 RECs in the UK. Clinical trial protocols were variable in the reporting of a range of items that are central to good trial conduct. Furthermore, this analysis did not reveal clear associations between general study characteristics and the reporting of methodological items.

With respect to the composition of the ASPIRE cohort, two key issues warrant consideration: the degree of missingness in the data and the generalizability of the cohort. As for missingness, this analysis demonstrated that only 35 of 288 eligible applications for clinical trials (12%) had missing protocols. Applications with missing protocols were also distributed among 12 of 19 RECs. Although the general characteristics of applications with missing protocols were different from applications that included a protocol, the reason underlying the missingness is also important. After approval was granted to allow access to the NRES database, it was quickly apparent that the NRES lacked the infrastructure to readily identify clinical trial protocols. At the time of our analysis, the NRES was transitioning from a mixed online and paper system for processing ethics applications to a completely online system. For trials in the ASPIRE cohort, the research documents corresponding to each application were only partially available online. Furthermore, a search for missing protocols was also unsuccessful and the RECs could not readily contact investigators to provide replacements. Hence, it is unlikely that the missingness in the ASPIRE cohort is related to the underlying characteristics of the study.

With respect to generalizability, findings based on the ASPIRE cohort are enhanced if the trials reflect broader clinical trial conduct and can be used to inform the development of international guidelines. Generalizability was first assessed by comparing the characteristics of studies approved by the Bristol Regional office to studies approved by other regional offices in the UK in 2012. Notably, the Bristol Regional office approved 23% of all studies in the UK and the percentage of CTIMP, phase I and phase II-IV trials were similar between the Bristol Regional office and non-Bristol offices. Although a more detailed characterisation of the studies approved by each regional office could not be performed, it is unlikely that there would be marked differences. Indeed, the most important determinants of where applications for ethics approval are submitted are proximity of the investigator to the REC and the availability of the REC to review the application. Undoubtedly, there will be variation in the mix of clinical areas addressed in trials across the 5 regions, reflecting research expertise in particular centres. For instance, the Oxford REC may receive a higher number of phase 1 trials of novel vaccines because there is a major vaccine research group located in its jurisdictional area. However, it should be noted that the five regional offices will nonetheless process applications from all clinical areas. Differences in the case mix of studies would therefore be minimal and unlikely to dramatically alter the major conclusions of this thesis.

Specific to the ASPIRE cohort, this analysis demonstrated that the trials in the cohort were large multicentre trials. Furthermore, many of the single centre trials were phase I trials, which were conducted in contract research organizations on behalf of international sponsors. Findings based on analyses of the ASPIRE cohort may therefore be generalizable to other jurisdictions outside of the UK.

Another important analysis in this chapter was the assessment of the quality of clinical trial protocols using the SPIRIT checklist. This analysis is the first to assess a cohort of clinical trials, irrespective of medical specialty, funding source and intervention type. The findings demonstrated that prior to the publication of the SPIRIT checklist, the quality of reporting of clinical trial protocols was modest. While some SPIRIT items were frequently addressed, other items – many of which are needed to reduce bias and have been a key area of advocacy by scientific journals and guideline committees – were not addressed. These findings are consistent with a similar assessment of 75 clinical trial protocols approved by the UK National Institute for Health Research (NIHR) from inception to August 2013 (12). Although these authors did not examine as many SPIRIT items as the analysis within this thesis, important parallels can be identified. For instance, only 50% (95% CI 39-62) of protocols in their study provided details on the criteria for discontinuing or modifying allocated interventions as well as strategies to improve adherence to the intervention (12). The corresponding proportions in this thesis' analysis were 55% (95% CI 44%-65%) and 42% (95% CI 31%-54%) respectively. Likewise, only 35% (95% CI 25%-46%) of trial protocols in the analysis of NIHR trials provided details on who will enrol/assign participants to ensure allocation concealment is maintained (12). In this thesis, the corresponding proportion was 33% (95% CI 26%-43%). Invariably, there will be some SPIRIT items that are better reported in the NIHR analysis compared to the ASPIRE cohort and vice versa. However, there remain important deficiencies in the reporting of important methodological items in clinical trial protocols.

Implications for Clinicians, Policy Makers and Future Research

Given variability in the quality of reporting of clinical trial protocols, the SPIRIT guidelines have the potential to improve the design and planning of clinical trials. However, the impact of SPIRIT will be contingent on whether it is embraced by RECs and researchers. The results of the current analysis provide empirical evidence of areas for improvement and can be used in advocacy to encourage adoption of SPIRIT. Given that the ASPIRE study is conducted with the support of the UK NRES, the adoption of the SPIRIT guidelines in the UK may be readily achievable.

Another strategy to improve the use of the SPIRIT guidelines is through the development of a writing tool that provides structured templates for clinical trial protocols based on the SPIRIT guideline. The development of such a tool is underway and the prototype is known as SEPTRE. Existing evidence based on the utility of other writing tools suggests that such efforts may yield benefit. For instance, investigators designed an RCT to assess the impact of a writing aid tool that was based on the CONSORT guidelines (14). Participants were randomized to the use and non-use of the tool to write the methods sections for a simulated RCT (14). Based on a global completeness of reporting score, use of the tool was associated with high completeness of reporting of the methods sections (mean difference 2.1, 1.5-2.7) (14). Whether the SEPTRE tool may yield similar benefits for the preparation of clinical trial protocols warrants further assessment (14).

Strengths and Limitations

The primary limitation of this analysis is related to the effort to condense the 270 item SPIRIT checklist to 83 items. While this task was done through independent efforts by two investigators with further consensus by two additional investigators, there may be

items that were inappropriately excluded during the extraction. Furthermore, extraction for SPIRIT items was limited to reports of non-pilot RCTs. Whether the reporting in pilot studies and NRTs is of similar quality warrants further study.

Conclusion

In summary, the ASPIRE cohort is made up of a wide range of clinical trials and the findings generated from the cohort are likely to be generalizable to the conduct of clinical trials in other jurisdictions. Furthermore, the quality of reporting of trial protocols is modest.

4

RESULTS

THEME: SELECTION OF OUTCOMES

Assessing the Use of Patient Reported Outcome Measures in Protocols for Clinical Trials

PROMs are reports obtained directly from patients with respect to their health and wellbeing. PROMs can relate to symptoms, signs, or other aspects of health such as convenience. PROMs are important outcomes in clinical trials that intend to demonstrate an improvement in participants' symptoms (1,2). They are also particularly important in clinical trials of chronic diseases wherein participants may require lifelong use of therapeutic agents (1,2).

Although the use of PROMs in clinical trials has been encouraged (3,4), recent estimates suggest that they may not be widely embraced. For instance, between 1990 and 2010, the use of PROMs increased from 2% to 15% in published cardiovascular trials (2).

Among trials registered in clinicaltrials.gov, the use of PROMs only increased from 25% to 30% between 2007 and 2013 (5). Furthermore, PROMs have not played an important role in the drug approval process, despite encouragement from regulatory agencies (3,6). To date, drugs with labelling claims based on PROMs – e.g. drug X improves pain – are rare. For instance, investigators assessed documents supporting the approval of novel agents from 1997 to 2005 (7). This was prior to the release of FDA draft guidance on the use of PROMs in 2006. Among 835 drug approvals, 194 (23%) included labelling claims related to PROMs (7). A subsequent study examined FDA documents supporting drug approvals from 2006 to 2010 and these authors found that among 116 drug approvals, 28 (24%) were granted labelling claims related to PROMs (1). Notably, 24 of these 28 labelling claims were for symptom relief, most commonly pain relief (1). Finally, the most updated analysis of FDA documents for PROM-related labelling claims was based on 40 FDA drug approvals for novel hematology and oncology drugs between 2009 and 2014 (8). In this small subset, only three (8%) received PROM related labelling (8).

There are many reasons why PROMs may be underused in clinical trials, despite encouragements from the FDA and EMA. First, there is inconsistency in the thresholds required by different regulatory agencies and this makes the regulatory landscape difficult to navigate. For instance, investigators compared labelling claims approved by the EMA and the FDA between 2006 and 2010 (9). For 75 products approved by both agencies, 25 (47%) had an EMA labelling claim related to PROMs whereas only 14 (19%) had an FDA labelling claim related to PROMs (9). When the two agencies provided a labelling claim for a drug, there was good concordance in the specifications of the labels (9). The differences between the agencies primarily reflected disagreement on what health concepts can be readily

assessed using PROMs (9). For instance, the EMA may approve claims of improved physical functioning based on subsets of the health related quality of life measure but the FDA would not accept this measure (9). Similarly, the EMA readily accepts PROMs that quantify fatigue whereas the FDA does not (9).

A second reason for the underuse of PROMs is that when the expectations from the EMA and FDA were consistent, they were nonetheless very stringent (3,6,10). For instance, PROMs that are submitted to support drug labelling must demonstrate content validity, which is whether the PROM actually measures what it is intended to measure (11). Content validity is important in a regulatory setting because the drug labelling must be a clear and accurate reflection of the data submitted (11). If PROMs are to be adapted for use in a new setting, qualitative and quantitative evidence must be provided to demonstrate the validity of the PROM (11). Furthermore, patient input must also be considered in the qualitative validation of the PROM and there must be a quantitative assessment to establish clinically meaningful thresholds for its use in a new setting (11). Content validity of PROMs has been a particular concern for drug labelling claims submitted to the FDA. For instance, the Transitional Dyspnea Index (TDI) was developed as a diagnostic tool for dyspnea and it was adapted for use in a clinical trial of an inhaler for adults with chronic obstructive pulmonary disease (11). However, the TDI was developed based on patient recall in short-term clinical trials and investigators in the inhaler trial adapted the TDI for the assessment of dyspnea at 1 year (11). The discrepancy in the timing of ascertainment and the lack of validation of the TDI for a 1-year endpoint was a contributing factor in the rejection of the patient reported outcome labelling claim (11). Notably, the EMA does not have specific requirements related to the timeline for patient recall (12).

The FDA and EMA guidance also provides specific recommendations related to the clinical trial design. These include pre-specification of PROMs, justification of the frequency and timing of the assessment of PROMs, careful consideration of the mode of collection (electronic versus paper) and clearly detailed methods to handle missing data and multiple testing (11). Missing data is particularly important and has been a key reason labelling claims based on PROMs have been rejected by the FDA. For instance, in a trial of an intervention to improve hypoactive sexual disorder, the investigators used PROMs based on patient diaries to measure any change in participant sexual desire (11). However, the clinical trial included strict criteria for compliance, which led to ≥25% participant drop out (11). There was also differential missingness in the intervention and placebo group, suggesting that missingness was related to poor treatment benefit (11). Finally, the investigators employed ad-hoc approaches such as complete case analyses to deal with missingness because the magnitude of missingness was not anticipated (11). These approaches were considered potentially biased by the FDA and the labelling claim was rejected (11). Missing data is therefore a particularly unique challenge when PROMs are used and strategies to deal with this issue require careful planning.

Although the FDA and EMA guidance apply only to trials that are intended for submission as part of the drug approval process, the thresholds put forward in their guidance documents often serve as a standard for well conducted research. This may also affect the willingness of other trialists to include PROMs in studies that are not submitted to these agencies.

The slow uptake of PROMs as well as the strict requirements for their use in clinical trials suggests that a more detailed analysis on how PROMs are incorporated in

contemporary clinical trials is warranted. To date, estimates of the use of PROMs have been derived from published trials or clinical trial registry entries, both of which have important limitations. With respect to published trials, these documents are representative of the research regulatory environment at the time of their inception, which may be up to ten years before the trial is published. There is also a risk of selective outcome reporting (13) and non-publication (14,15), which undermines estimates based on published trials. With respect to trial registries, the fidelity and comprehensiveness of information included in trial registries has not been assessed (16) and it is unclear whether all study outcomes are consistently included in the registry entry. Finally, there are no studies that have specifically examined methodological issues related to the use of PROMs. Analyses of trial protocols, such as the ASPIRE cohort, overcome these limitations and may provide contemporary and meaningful insight into the use of PROMs.

The co-primary objectives of this study were to determine what proportion of clinical trial protocols included PROMs and to assess the inclusion of PROMs based on trial characteristics such as the type of outcome and intervention. Second, this study sought to determine whether methodological challenges related to the use of PROMs – validity, sample size calculations and missing data – were accounted for in trial protocols.

4.1 CHAPTER SPECIFIC METHODS

Definitions

PROMs: The definition of PROMs provided by the United States FDA was used. PROMs are reports coming directly from patients about how they feel or function without interpretation by healthcare professionals or anyone else (3). PROMs can relate to symptoms, signs, functional status, perceptions, or other aspects such as convenience and tolerability (3). Assessments of adverse events were not included as PROMs because it was expected that all trials should include details on adverse events. Examples of PROMs include health-related quality of life scales, disease severity questionnaires such as the asthma control questionnaire and visual analogue scales (See Appendix 2 for further examples).

Inclusion and Exclusion Criteria

This study was limited to phase III RCTs in the ASPIRE cohort because these trials are intended to change clinical practice and PROMs will be most relevant in this setting (2).

Data Extraction

Information on the general characteristics of the trials was extracted, as described in Chapter Three. All outcomes for protocols of RCTs in the ASPIRE cohort were assessed to determine if PROMs were included and whether this was as a primary, secondary or "other" outcome. The primary outcomes of trials were also classified as clinical outcomes or surrogate/physiology outcomes. Clinical outcomes were defined as outcomes that directly pertained to a patient's health or well-being (17). These included PROMs as well as death and events that adversely affect quality of life (major morbid events such as stroke and minor morbid events such as infection or pain) (17). Surrogate outcomes were defined as intermediate end points focused on disease progression or serving as an indicator for

subsequent risk (17). These included outcomes such as blood pressure, serum cholesterol, and progression free survival. Physiological outcomes were outcomes which focused on laboratory changes without a direct effect on patient functioning.

Trials were also classified by medical specialty, by the design of the trial, and based on whether participants were blinded during the study. The intervention used in the trial was classified as pharmacologic, behavioural or device/procedure/surgery. The duration of the intervention was classified based on clinical judgement. It was acute if the expected use of the intervention was less than one month or intermediate/chronic if the expected use was more than one month (18). The number of centres involved in the trial was extracted and the funding source of the trial was classified as industry versus non-industry funded. Finally, using the IRAS form, the trials were classified based on whether the study was intended for submission to regulatory authorities. The use of PROMs were compared based on the type of outcomes (clinical versus surrogate), intervention type (pharmacologic versus non-pharmacologic), length of intervention (acute versus intermediate/chronic), funding source (non-industry versus industry funding) and whether the trial was intended for submission to regulatory authorities.

Studies with PROMs as Primary Outcomes

Based on recommendations included in the CONSORT extension for PROMs (19), trial protocols with PROMs as primary outcomes were assessed to determine if investigators provided details on the instrument validity of the PROMs, described methods for handling missing data for PROMs and reported sample size calculations based on the PROMs. The

latter two items are important for studies with PROMs as primary outcomes because the participant burden associated with PROMs results in an increased potential for missing data and loss to follow-up.

Consistent with the CONSORT extension for PROMs (19), a liberal criterion was used for the reporting of the validity of PROMs. Any attempt to discuss the robustness of the PROMs included in the trial or any reference to a validation study was deemed sufficient. With respect to missing data, each protocol was assessed for whether investigators provided details on strategies for handling missing data, including but not limited to multiple imputation. The quality of the missing data management strategy was not assessed. Therefore, studies using last observation carried forward were also noted as providing details about missing data.

Finally, the assessment of sample size calculations in PROMs was two-fold. First, studies were assessed to determine if details pertaining to the core components of sample size calculations were provided (20). This included details of the minimal clinical important effect size (delta), established event rates in each study arm (if binary outcome), the standard deviation for effect size (if continuous outcome), type I error, type II error and the anticipated loss to follow up.

Using the estimates provided in the study for these core components, the sample size calculation in each RCT with a PROM as the primary outcome was replicated. In the event that the study did not report the type I error or type II error, the sample size was replicated using a two tailed 5% type I error and 20% type II error (21,22). If the anticipated number of drop outs was not reported, the sample size was calculated without any additional adjustment (21,22). Protocols that did not report the minimally important difference, standard deviation

or event rate were classified as having sample sizes that could not be replicated. Sample size calculations were replicated using PASS 11. For a binary end point, the formula for a chi-squared test was used (unless a fisher's exact test was specified in the trial protocol). For a continuous end point, the formula for a Student's t-test was used and for a time to event end point, the formula for a long-rank test was used. If the replicated sample size was within 5% of the sample size reported in the protocol, then it was considered reproduced. If the sample size was not within 5%, the sample size calculation was repeated with Stata and R. If the independent replication was not within 5% of the sample size reported in the protocol, then the sample size was not considered reproducible. Where studies used simulation to calculate the sample size, the study was excluded (n=4).

Statistics

A descriptive analysis was performed, as detailed in the Methods Chapter. The use of PROMs was assessed using multilevel poisson regression analyses with robust standard errors. The potential predictors of interest in multivariable models were the type of outcome (surrogate/physiological versus clinical), intervention type (pharmacologic versus non-pharmacologic), the duration of intervention (acute versus chronic intervention duration), funding source (industry versus non-industry funded), and whether the trial was intended for submission to regulatory authorities. Models were also adjusted for whether the study participants were blinded, disease area (cardiovascular versus non-cardiovascular trial) and sample size (continuous). Sample sizes reported in the protocol were compared to the replicated sample sizes using Bland-Altman plots.

4.2 CHAPTER RESULTS

General Characteristics of Included Trials

The study cohort was composed of 132 phase III RCTs (Figure 4.1) and the general characteristics are detailed in Table 4.1. Most studies examined medical conditions related to oncology, psychiatry, autoimmune/musculoskeletal conditions or cardiovascular disease. Fifty-one of 132 RCTs (39%) involved a pharmacologic intervention and 87 (66%) involved intermediate or chronic interventions. Thirty RCTs (23%) were multinational and 78 (59%) were multicentre trials. Finally, 46 (35%) were industry funded and 40 (30%) were intended for submission to regulatory authorities. The general characteristics of studies that included PROMs as primary outcomes are also detailed in Table 4.1.

Figure 4.1: Identification of Studies Included in the Analysis of Patient Reported Outcome Measures

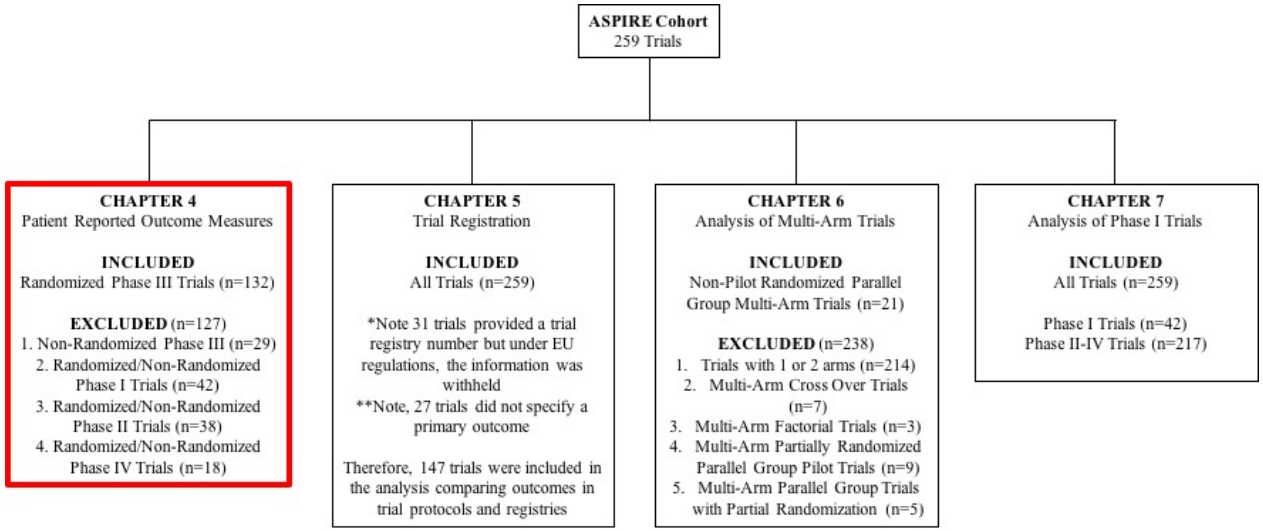


Table 4.1: General Characteristics of Included Randomized Controlled Trials

	Number of Trials (%)	
	Overall (n=132)	**PROMs as Primary Outcome (n=44 of 132)**
Type of Outcome*		
Surrogate/Physiological Outcome	62 / 88 (70)	-
Clinical Outcome	26 / 88 (30)	-
Disease Area		
Non-Cardiovascular Disease	111 (84)	41 (93)
Cancer	20 (15)	3 (7)
Dermatologic	1 (1)	0 (0)
Gastroenterology	6 (5)	1 (2)
Musculoskeletal and Autoimmune Disease	15 (11)	8 (18)
Neurologic	10 (8)	3 (7)
Obstetrics	5 (4)	1 (2)
Opthamology	5 (4)	1 (2)
Psychiatry	17 (13)	12 (27)
Respiratory Medicine	9 (7)	3 (7)
Other	23 (17)	15 (34)
Cardiovascular Disease**	21 (16)	3 (7)
Design		
Parallel Group	118 (89)	43 (98)
Crossover	7 (5)	1 (2)
Other	7 (5)	0 (0)
Intervention (%)		
Drug	51 (39)	6 (14)
Behaviour/Lifestyle	32 (24)	21 (48)
Procedure/Surgery	20 (15)	9 (21)
Biological/Vaccine	4 (3)	0 (0)
Device	15 (11)	6 (14)
Dietary Supplement	4 (3)	2 (5)
Other	6 (5)	0 (0)
Length of Intervention		
Acute	45 (34)	14 (32)
Intermediate/Chronic	87 (66)	30 (68)
Multinational		
Yes	30 (23)	0 (0)
No	102 (77)	44 (100)
Number of Centres		
Single	53 (41)	22 (50)
Multiple	78 (59)	22 (50)
Unclear	1 (1)	0 (0)

Number of Arms		
2	106 (80)	36 (82)
3	19 (14)	7 (16)
4	5 (4)	1 (2)
>4	2 (2)	0 (0)
Sample Size	106 (51-310)	382 (82-163)
Funding		
Non-Industry	86 (65)	41 (93)
Solely or Part Industry	46 (35)	3 (7)
Submission to Regulatory Authorities		
Yes	92 (70)	41 (93)
No	40 (30)	3 (7)

*Studies in which PROMs were the primary outcome were excluded **Cardiovascular disease trials include trials directly focusing on cardiovascular disease as well as trial focused on cardiovascular risk factors: diabetes mellitus, hyperlipidemia, and chronic kidney disease

Ninety-four of 132 RCTs (71%) included PROMs as a study outcome. Studies examining chronic interventions were more likely to include PROMs (adjusted relative risk [aRR]: 1.44, 1.02-2.03, Table 4.2). However, this estimate was modest and the confidence intervals were close to the line of no effect. There was no association between industry funding and the use of PROMs (aRR 0.90, 0.64-1.25, Table 4.2).

Table 4.2: Inclusion of Patient Reported Outcome Measures in Protocols of Randomized Controlled Trials

	Number of Trials (%)					
	PROs Included (n=94)	**PROs Not Included (n=38)**	**Unadjusted RR (95% CI)**	**P-value**	**Adjusted RR (95% CI)*****	**P-value**
Type of Outcome (n=88)*						
Surrogate/Physiological Outcome	33 (53)	29 (47)	ref	ref	ref	ref
Clinical Outcome	17 (65)	9 (35)	1.23 (0.87-1.74)	0.25	0.97 (0.68-1.37)	0.859
Intervention (%)						
Non-pharmacologic	56 (77)	17 (23)	ref	ref	ref	ref
Pharmacologic or Vaccine	38 (64)	21 (36)	0.84 (0.68-1.04)	0.161	0.85 (0.61-1.19)	0.34
Length of Intervention						
Acute	27 (60)	18 (40)	ref	ref	ref	ref
Intermediate/Chronic	67 (77)	20 (23)	1.28 (0.95-1.73)	0.11	1.44 (1.02-2.03)	0.040
Funding						
Non-Industry	64 (74)	22 (26)	ref	ref	ref	ref
Solely or Part Industry	30 (65)	16 (35)	0.88 (0.73-1.05)	0.161	0.90 (0.64-1.25)	0.52
Submission to Regulatory Authorities						
No	66 (72)	26 (28)	ref	ref	ref	ref
Yes	28 (70)	12 (30)	0.98 (0.79-1.20)	0.82	0.92 (0.68-1.25)	0.60

"ref" is references and RR is relative risk. Proportions are compiled across rows. *Studies in which PROMs were the primary outcome were excluded ***Adjusted for whether the study participants were blinded (all multivariable models), disease area (cardiovascular versus non-cardiovascular trial) and sample size (continuous).

Reporting of Methodological Items

Among 94 RCT protocols that included PROMs, 44 (47%) studies included PROMs as primary outcomes. Eighteen of these 44 RCTs (41%) provided details about the validity of the PROM used in their trial and only 10 (22%) RCTs provided details about how missing data would be handled (Table 4.3). With respect to sample size calculations, six RCTs explicitly stated that a sample size calculation was not performed because the study was a pilot trial (Table 4.3). Of the remaining 38 RCTs, only 13 (34%) provided details about all the core components of the sample size calculation (Table 4.3). Notably, only 17 (44%) RCTs provided the anticipated loss to follow up used for the derivation of their sample size calculation.

Table 4.3: Methodological Items Related to the Use of Patient Reported Outcome Measures as a Primary Outcome in Clinical Trials

Methodological Items Reported	**PROMs as Primary Outcome (n=44) n (%) [95% CI]**
Validity of PROMs	18 (41) [27-56]
Strategies for Missing Data	10 (22) [12-38]
Sample Size Calculation*	**(n=38)**
All components reported	13 (34) [20-51]
Minimal clinical important effect size (delta)	24 (63) [46-77]
Established event rate in each study arm (if binary outcome)	3 / 4 (75) [4-100]
Standard deviation for effect size (if continuous outcome)	18 / 34 (53) [36-70]
Type I error	24 (63)

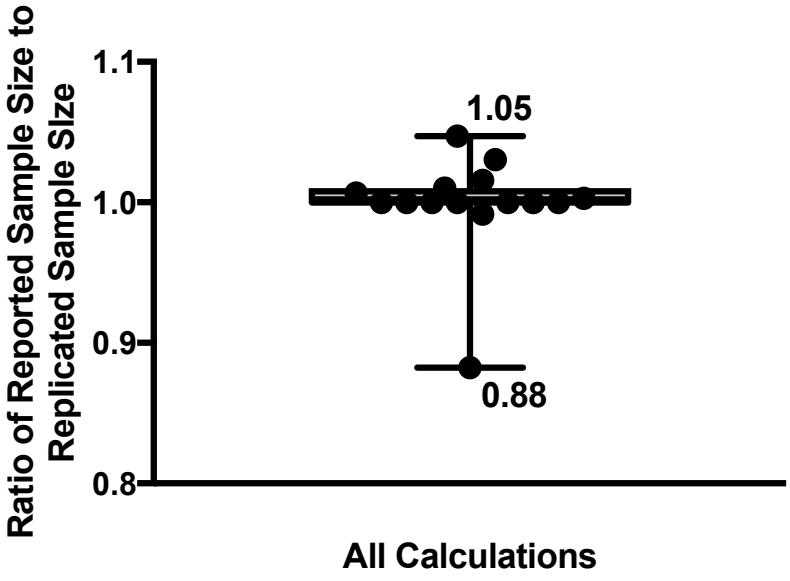
	[46-77]
Power	28 (74) [57-86]
Anticipated loss to follow up	17 (44) [29-61]

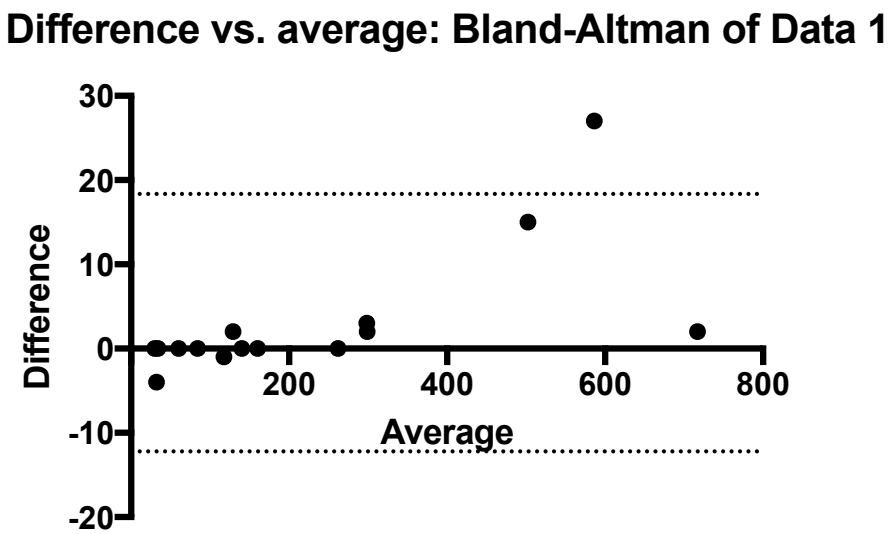
*Six studies explicitly indicated that a sample size calculation was not performed.

Replication of Sample Size Calculation

There were 19 trials for which a sample size replication could be performed. In four trials the sample size could not be replicated because simulations were used. For the remaining 15 trials, the ratio of the reported and replicated sample size as well as the agreement between the two estimates are displayed in Figure 4.2. The reported sample size was could not be replicated for one trial (ratio of sample sizes 0.88). There was otherwise good agreement between the two estimates (Bias 3.1, -12 to 18).

Figure 4.2: Comparison of Reported and Replicated Sample Sizes





4.3 CHAPTER DISCUSSION

This study is a contemporary analysis of the use of PROMs in clinical trials. The major findings were two-fold. First, PROMs were used in 71% of phase III RCTs receiving research ethics approval in the UK; Second, methodological items related to the use of PROMs as primary outcomes were infrequently reported.

Despite encouragement for the inclusion of PROMs as primary and/or secondary in clinical trials (4,23), recent studies suggest that these outcomes are used in 15% to 30% of trials (2,5). However, these estimates were derived from assessments of published trial reports and clinical trial registries, both of which may provide incomplete details about the outcomes included in a trial. This study thesis to directly address this limitation by focusing on a contemporary cohort of clinical trial protocols. Furthermore, there was a specific focus on phase III trials, which are designed to inform clinical practice and may therefore be more likely to include PROMs. In the current analysis, 71% of RCTs included PROMs.

However, this thesis did not clearly identify any study characteristics associated with underuse of PROMs. Although some relative risk estimates achieved p-values that would be considered statistically significant, the confidence intervals were close to the line of no effect and could easily be attributed to residual confounding. Accordingly, this thesis suggests that that PROMs are broadly embraced in phase III trials, which is the context in which their use would be considered most relevant.

This study also noted that some key methodological items that are pertinent when PROMs are included as a primary outcome – validity of PROMs, strategies for missing data

and components of sample size calculations – were infrequently addressed in clinical trial protocols. This oversight has the potential to yield biased estimates. First, the validity of PROMs is central to their use in clinical trials because it is important to ensure that PROMs actually measure the construct of interest. Indeed, when the validity of PROMs is questionable, it is unclear whether the results from the PROMs are indeed a true reflection of patient functioning or symptoms (1). Second, failure to apply robust methods for addressing missing data is particularly problematic (24). For instance, in pharmacological trials, participants who experience toxicity may be less likely to complete questionnaires or other PROMs. If missing data from these participants is not adequately accounted for, this would lead to an underestimation of the adverse effect of an intervention on the patient's functioning and erroneous recommendation of interventions that may be harmful (25).

Finally, details regarding sample size calculations were often absent in trial protocols. Although sample size calculations are relevant in all trials, they are particularly important in trials where PROMs are chosen as the primary end-point due to the potential for loss to follow-up. Notably, only 17 (44%) of 38 RCTs in this analysis provided the anticipated loss to follow up used for the derivation of their sample size calculation. For the 15 trials where a sample size replication could be performed, there was good agreement between the reported sample size and the replicated sample size. There was only one instance where the reported sample size underestimated the replicated sample size. Gaps in the reporting of these methodological details may undermine the findings of studies which use PROMs as a primary end point.

Implications for Clinicians, Policy Makers and Future Research

Improvements in the planning and use of PROMs in clinical trials may be achieved through the development of a single consolidated guideline on PROMs. A recent systematic review assessed the quality of guidance for the use of PROMs in trials (26) and the study identified 8 published guidance documents and 46 grey literature documents. Together, these documents provided 162 recommendations about what should be included in clinical trial protocols involving PROMs (26). However, only 5 recommendations consistently appeared in each of the guidance documents, demonstrating the lack of consistency among these resources (26). As of July 2017, the SPIRIT initiative does not currently provide guidance on the use of PROMs and may therefore serve as an important and centralized venue to inform researchers about the inclusion of PROMs in clinical trials.

The appropriate use of PROMs in clinical trials can also be ensured during the research ethics review process. For instance, funding bodies as well as RECs should ensure that trials incorporate PROMs that are well validated, with adequate consideration of important statistical and methodological issues. In recognition of the difficulty associated with developing robust PROMs, the FDA works collaboratively with trialists that are including PROMs in their study (11). However, there is no comparable support available for researchers conducting non-industry funded clinical trials. If RECs or funders provide a similar centralized support, this may increase value from research based on PROMs.

Strengths and Limitations

Some limitations should be considered when interpreting the results of this study. First, this thesis chapter focused on phase III trials, but PROMs may be used in other trials.

However, phase III trials are intended to guide clinical practice and the assessment of PROMs is these studies will have most relevance to the interventions that are ultimately used. Second, this study did not assess the appropriateness of strategies to manage missing data or assumptions of the sample size calculation. However, the reporting of these elements was already particularly poor and a further detailed assessment would not enhance this analysis.

Conclusion

PROMs are not consistently used across all clinical areas and methodological items related to PROMs are not consistently reported in trial protocols. Including PROM specific details in guidelines for protocol preparation may improve how PROMs are incorporated into clinical trials.

5

RESULTS

THEME: TRANSPARENCY AND ACCOUNTABILITY

Assessing Concordance of Primary and Secondary Outcomes in Clinical Trial Protocols and Trial Registries

Non-publication of clinical trials has become a well publicized issue, in large part because of the suppression of trial results related to adverse events (1,2). As a result, there is a bias in the scientific literature, favoring the beneficial results of new treatments (1,2). For instance, one of the most well publicized cases of non-publication related to clinical trials of paroxetine, an antidepressant, in children with major depressive disorder (3). Clinicians and policy makers were concerned that use of these medications increased the risk of suicide but evidence based on published trials suggested that paroxetine was associated with benefit (3). However, in a seminal analysis, investigators were able to gain access to unpublished studies on antidepressant use in children and demonstrated that once unpublished studies were

included in a meta-analysis, the risks of anti-depressants outweighed the benefits (4). This resulted in the revision of guidelines supporting the use of paroxetine in children with depression and also formed the basis of a legal challenge against Glaxo-Smith-Kline, the manufacturer of paroxetine (3).

Similarly, antiviral agents such as oseltamvir are recommended for the treatment of influenza (5). Antivirals are also actively stockpiled by government agencies in anticipation of a large and poorly controlled influenza outbreak (5). However, data on the harms of antivirals remained unpublished (5). When investigators gained access to clinical study reports – which are extensive summaries of RCTs of drugs – the authors demonstrated that oseltamivir provided modest symptom reduction in adults with influenza but also caused nausea, vomiting, headaches as well as renal and psychiatric complications (5).

Clinical trial registration was introduced as a strategy to reduce non-publication of trials and to improve the transparent and accurate reporting of research. The rationale underlying trial registration is that when investigators pre-specify the design, conduct, and analysis of their study in a publicly accessible database, there is increased accountability for conducting the study according to pre-specified plans. Furthermore, researchers and the general public have access to a comprehensive list of trials that are initiated and can compare the pre-specified details in the registry entry to the final publication of the trial (6,7).

There are already tangible benefits derived from clinical trial registration, particularly with respect to tracking non-publication of trials. For instance, an early analysis of the publication of trials registered in clinicaltrials.gov – a trial registry maintained by the US National Library of Medicine – demonstrated that only 46% (311 of 677) of trials that were registered and completed by 2005 were ultimately published (8). Notably, industry funded

trials were less likely to be published than non-industry/non-government funded trials (40% versus 56% respectively, $p<0.001$), but there was no difference between industry and government funded trials (40% versus 47% respectively, $p=0.22$) (8). Through analyses of trial registries, non-publication has also been shown to be particularly pervasive among trials funded by institutions such as the National Institute of Health in the US (9), trials conducted at academic medical centres (10) and also large clinical trials (11), which are particularly important for guiding the use of medical interventions.

Trial registries have also allowed researchers to assess for selective reporting of outcomes in clinical trials. For instance, investigators compared the trial registry entry and publication for 147 trials that were registered in clinicaltrials.gov (12). The authors found that 46 trials had different primary outcomes in the trial registry compared to the trial publication, such as the omission of a registered outcome in the trial publication, the addition of a new outcome in the publication that was not in the registry, changes in the priority given to the outcome (primary versus secondary) and changes in the timing of the assessment of the outcome (12). Notably, discrepancies in the primary outcome between the trial registry and publication were more likely to favor statistically significant results (12).

More recently, researchers have also used trial registries as a tool for advocacy, by contacting journals in which discrepancies were discovered between outcomes in study protocols or trial registries and published articles. This effort, known as the COMPare project, examined 67 trials, of which there were only nine where the pre-specified outcomes perfectly matched between the protocol/registry and the published article (13). Whereas some journals have readily corrected the scientific record once informed about the discrepancies by the COMPare project, others have made no changes (13).

Despite these established gains, the value of trial registration will only be maximized when all trials are registered before study initiation. The World Health Organization (WHO) (14), International Committee of Medical Journal Editors (ICMJE) (15), the US FDA (16) and the EMA have introduced policies to encourage trial registration. This combined effort has led to gradual improvement in the frequency of trial registration. For instance, of 616 RCTs published in PubMed indexed English language journals in December 2006, only 9% were reported as registered (17). A subsequent analysis of 614 RCTs published in 121 PubMed indexed English language journals in November 2010 found that 55% were reported as registered (18). While these finding are encouraging, published trials do not reflect the current research regulatory environment and may not adequately indicate the overall adoption of trial registration policies. As well, these estimates of trial registration may be incorrect if unpublished and/or discontinued trials are not included. An analysis of more contemporary trial documentation such as clinical trial protocols is required to obtain more robust estimates of trial registration.

Achieving the full value of trial registration is also contingent on the accurate and complete reporting of information about clinical trials in the registry (19). If this does not occur, comparisons between trial registries and publications may lead to inaccurate conclusions regarding what occurred during the conduct of the trial and whether it aligned with the apriori intentions of investigators. While many studies have compared trial registries to published articles (6,7,12), no study has compared clinical trial protocols to their registry entries. This step would be critical to assess the accuracy of information detailed within trial registries and provide insight into their utility.

The primary objectives of this study were therefore three-fold. First, it was to assess the frequency of trial registration. Second, it was to assess the fidelity of information in trial registries by comparing the primary and secondary outcomes in clinical trial protocols to the trial registry entry. Third, it was to determine the reasons trials were not registered based on detailed documentation in the IRAS form.

5.1 CHAPTER SPECIFIC METHODS

Randomized and non-randomized clinical trials identified as part of the ASPIRE cohort were assessed to determine if they were registered in a clinical trial registry. This was done in two ways. First, as part of the ethics application, researchers may report trial registration numbers for their study. If the study was registered in a WHO primary registry, the study was identified as registered. Second, for clinical trials without a registry number in the ethics application, a systematic search of the WHO International Clinical Trials Registry Platform, the European register of clinical trials, clinicaltrials.gov and the International Standard Randomized Controlled Trial Number (ISRCTN) was conducted. As recommended by bibliographic studies (20), the search was conducted on the basic search interface for all registries using the following search strategy: The title of the protocol; the name of the principal investigator and a combination of keywords (e.g. the intervention name, health condition, trial identifier). The search for all trials was completed in April 2016.

Data Extraction for Study Outcomes

Information on the general characteristics of the trials was extracted, as described in the Methods Chapter. For the comparison between the protocol and registry, individual study outcomes were only extracted for trials that were registered and explicitly identified outcomes in their protocol as primary or secondary.

The SPIRIT guidelines requires that at least 4 components are defined for each study outcome: 1) the classification of the outcome as primary, secondary or exploratory; 2) the measurement variable, which is the data collected from study participants (e.g. systolic blood pressure); 3) the analytic metric, which is the format in which the measurement variable is being collected for analysis (e.g. change from baseline, final value or time to event); 4) the time point at which the outcome is measured for analysis. If measurement variables (e.g. systolic blood pressure) were measured at multiple time points (e.g. 3 months and 6 months), each time point was included in the study as a separate study outcome.

The classification of outcomes in the trial protocol and registry were assessed for reporting using a 4-point scale (Primary, Secondary, Other or Not Mentioned). “Not Mentioned” was used if an outcome was reported in the study protocol but not in the registry or vice versa.

With respect to the measurement variable, analytic metric and time point, outcomes were assessed using a 4-point scale (Reported, Not Reported, Not Applicable or Outcome Not Mentioned). For the assessment of the analytic metric, “Not Applicable” was used if the outcome in the protocol was a count outcome (e.g. proportion of participants who complete the study). For the assessment of time points, “Not Applicable” was used if the outcome was a time to event outcome or if the investigators indicated that it was an overall outcome that would be examined over the course of the study (e.g. overall survival). Finally, “Not

Mentioned" was used if an outcome was newly added in the registry and not reported in the study protocol, or vice versa. Comparisons between the reporting of outcomes in the study protocol and the trial registry were made using cross-tabulation.

For a random sample of 30 NRTs and/or RCTs (20% of trials where outcomes were compared between the protocol and registry), amendments submitted to the REC were examined for any changes in the study outcomes. If an amendment was reported, it was assessed to determine if any discrepancies between the trial protocol and trial registry were accounted for by the amendment. Details about study amendments were available from the REC in two forms. First, a summary of the amendment was provided within the online interface for all trials. This summary was taken from the IRAS amendment form that researchers must submit when requesting a study amendment. Second, if the study amendment involved a change to the trial protocol, an updated version of the protocol was provided to the REC.

This retrospective approach to assess the impact of trial amendments was taken for two reasons. First, as the trials in the ASPIRE cohort were ongoing when the work for this thesis chapter began, many of the amendments would not have been available. Second, as mentioned in the Methods Chapter, the RECs did not have a well-established system for ensuring that trial documents were readily available on their online platform. It was anticipated that this would be particularly problematic for amendments to trials in the ASPIRE cohort. For any given trial, there may be multiple amendments and therefore, multiple updated versions of the protocol.

The NRES would need to hire new personnel to upload any missing documentation to their server. Requests for documentation would therefore be handled on an "as-needed"

basis and there was no guarantee that all missing documentation could be obtained. Indeed, as demonstrated in Chapter One, there were 35 ethics applications for which the trial protocol that was approved by the REC could not be obtained. The retrospective approach of examining a subset of trials to determine if amendments accounted for the findings in this thesis was therefore considered pragmatic. It also reflects the limited infrastructure and limited human resources available to the NRES.

Reasons for Non Registration of Trials

As part of the application for research ethics approval in the UK, investigators are required to respond to the following question: "*Article 19 of the World Medical Association Declaration of Helsinki adopted in 2008 states that "every clinical trial must be registered on a publicly accessible database before recruitment of the first subject"; and the International Committee of Medical Journal Editors (ICMJE) will consider a clinical trial for publication only if it has been registered in an appropriate registry. Please see guidance for more information. Will the research be registered on a public database? Please give details, or justify if not registering the research.*" Based on responses to this question, the reasons investigators gave for trials that were not registered were assessed and analyzed inductively. Two investigators – one of which was the DPhil candidate – performed the coding and identification of themes. These results were refined through discussion with other study investigators and codes were grouped together to identify common themes. Discussions continued until consensus was reached on which themes encompassed the responses in the ethics applications.

Data Analysis

A descriptive analysis was performed, as detailed in the Methods chapter. The reporting of each of the four components of study outcomes was compared between the protocol and trial registry. Each component was assessed separately and the focus of the analysis was to assess concordance in the reporting of outcomes between the trial protocol and trial registry. Concordance was defined as the number of outcomes where the component under assessment – classification, measurement variable, time point or analytic metric – was defined in the trial protocol and reported exactly the same way in the registry. When the component of the the outcome was either not reported, unclear, or deemed not applicable based on the trial protocol, concordance could not be assessed and this outcome was not used to generate the estimate of concordance. This is consistent with the approach employed in previously published studies (12) comparing study outcomes between the trial registry and published article.

There were also two levels considered in this analysis. First, the unit of analysis was the individual study protocols and second, the unit of analysis was the individual outcomes.

Where the unit of analysis was the study protocol, the number of studies that had each outcome component classified as “Reported and Concordant” between the protocol and registry was determined. If at least one primary or secondary outcome had one discrepancy, the study was classified as “Registry Entry Not Concordant with Protocol”. This assessment was done separately for each of the four components of study outcomes.

Where the unit of analysis was the individual study outcomes, cross-tabulation was used to compare the reporting of study outcomes in study protocols and trial registries. Furthermore, a combined estimate to summarize reporting across all four estimates was not

generated. This is because components such the time point or analytic metric may not apply for certain outcomes and a "global" assessment of concordance would be driven by the classification and measurement variable.

With the level of analysis as the individual study outcome, multilevel Poisson regression with robust errors was also used to assess associations between general study characteristics and concordance between trial protocol and trial registry. The model accounted for clustering at the level of the REC and the level of the study. ICCs were provided for this analysis in Table 5.1 because two levels of clustering were used. Furthermore, the extent of clustering at the level of the study may be high and it is important to understand the relative gain from conducting the analysis based on individual study outcomes. The predictors of interest that were included in multivariable models were funding source (industry versus non-industry funded), intervention type (pharmacologic versus non-pharmacologic), the number of centres (multiple versus single centre), and sample size (continuous). Models were further adjusted for whether the outcome was primary, secondary or other, and for whether the study was a pilot study.

Table 5.1: Intraclass Correlation Coefficients and Selected Regression Models for Analysis of Discrepancies Between Trial Protocols and Trial Registries (Analysis is at the level of the study outcome)

Outcome	Cluster Variable	ICC	95% CI
Any Discrepancy	Research Ethics Committee	0.14	0.03-0.43
	Study	0.83	0.75-0.89
Variable Classification	Research Ethics Committee	0	0
	Study	0.75	0.68-0.82
Measurement Variable	Research Ethics Committee	0.05	0.005-0.38
	Study	0.79	0.72-0.85
Analytic Metric	Research Ethics Committee	0.15	0.04-0.46

	Study	0.80	0.72-0.87
Time Point	Research Ethics Committee	0.01	0-1
	Study	0.86	0.79-1

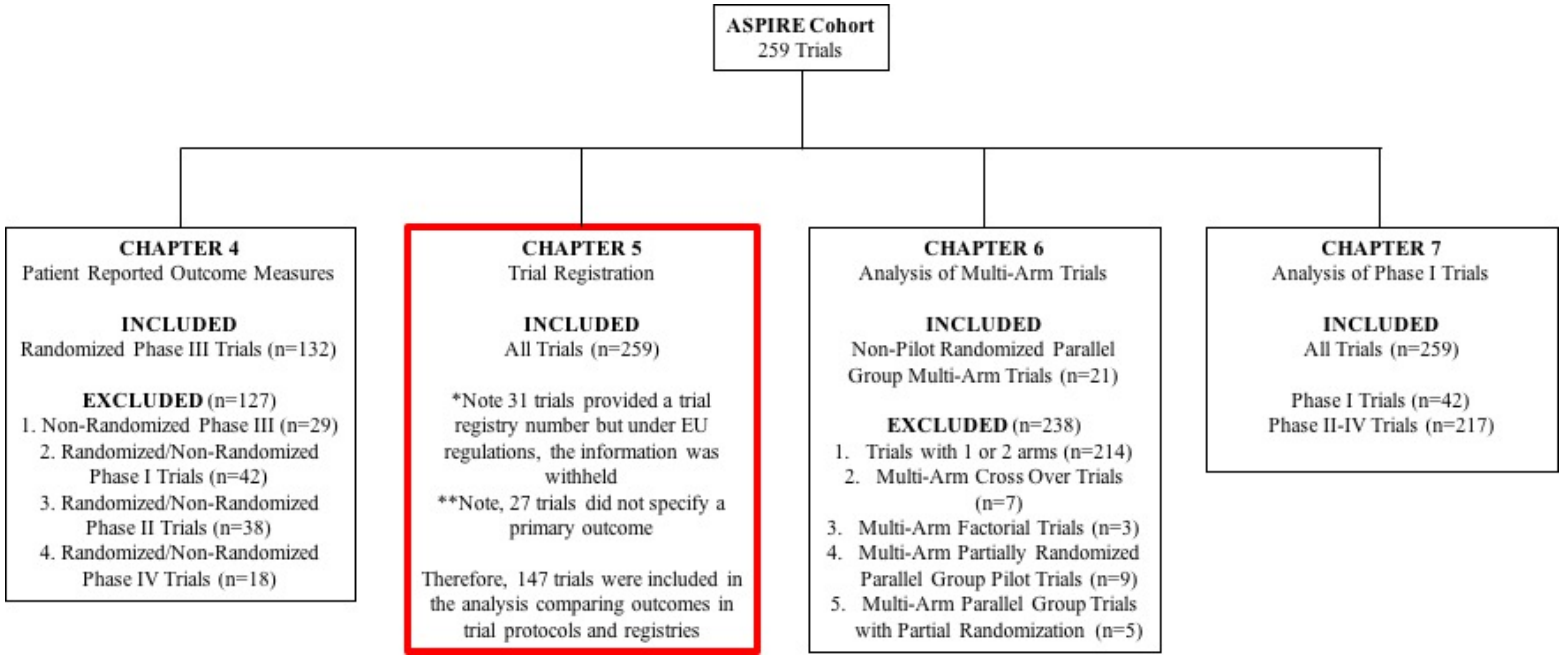
ICC is the intraclass correlation coefficient

5.2 CHAPTER RESULTS

General Characteristics of Included Trials

The study cohort was composed of 259 clinical trials (Figure 5.1), of which 205 (79%) were registered in a trial registry and 54 (21%) were not registered. The general characteristics are detailed in Table 5.2.

Figure 5.1: Identification of Included Studies in the Analysis of Trial Registration



Eighty-two percent of randomized trials were registered. Trial registration was common for pharmacologic interventions, procedures and surgeries, biological and vaccine agents, and dietary supplements but less frequent among trials involving behavior and lifestyle interventions, and devices (Table 5.2). Trial registration was also common among multicentre trials (88%) but less common among single centre trials (67%). Registered trials were larger in sample size. Finally, ninety-four percent of industry funded trials were registered and 64% of non-industry funded trials were registered.

Table 5.2: General Characteristics of Included Studies in the Analysis of Trial Registration

	Not Registered (n=54)	**Registered (n=205)**	**P-value**
Randomized (%)	34 (18)	159 (82)	0.035
Design (%)			
Parallel Group	28 (16)	144 (84)	0.002
Crossover	4 (22)	14 (78)	
Not Applicable (Single Arm Intervention Study)	19 (31)	42 (69)	
Other	3 (37)	5 (63)	
Intervention (%)			
Drug	8 (6)	131 (94)	<0.001
Behaviour/Lifestyle	21 (48)	23 (52)	
Procedure/Surgery	8 (33)	16 (67)	
Biological/Vaccine	3 (25)	9 (75)	
Device	13 (45)	16 (55)	
Dietary Supplement	0 (0)	4 (100)	
Other	1 (16)	6 (84)	
Number of Centres (%)			
Single	35 (33)	70 (67)	<0.001
Multiple	18 (12)	134 (88)	
Unclear	1 (50)	1 (50)	
Number of Arms (%)			
1	20 (30)	46 (70)	0.08
2	27 (18)	120 (82)	
3	6 (24)	19 (76)	
4	1 (8)	12 (92)	
>4	0 (0)	8 (100)	

Sample Size (Not Randomized)	18 (25-40)	45 (22-117)	0.027
Sample Size (Randomized)	60 (36-120)	100 (50-300)	0.008
Funding (%)			
Non-Industry	46 (36)	84 (64)	<0.001
Solely or Part Industry	8 (6)	121 (94)	

Proportions are computed across rows. Sample sizes are given as median (interquartile range).

Discrepancies Between Clinical Trial Protocols and Trial Registries: Analysis at the Level of Trial Protocol

Among 205 trials that were registered, 31 (15%) provided a trial registry number but under European Union regulations, the information was withheld in the trial registry. An additional 27 (13%) trials did not specify a primary outcome in the trial protocol. Accordingly, 147 trials were included in the analysis comparing study outcomes in the protocol to the trial registry.

With respect to primary outcomes, 132 of 145 trials (90%) listed all primary outcomes in the protocol as primary outcomes in the trial registry (Table 5.3). The measurement variable for the primary outcome in the protocol was concordant with the trial registry for 128 trials (87%, Table 5.3). In contrast, the analysis metric was concordant between the protocol and registry for 59 trials (58%). However, the analytic metric was not specified in the protocol for 16 trials (16%) and therefore concordance could not be assessed (Table 5.3). Finally, the time point of analysis was concordant between the protocol and registry for 90 trials (78%) but it was not specified in the protocol for 17 trials (17%, Table 5.3).

One hundred and forty-five trial protocols included at least one secondary outcome, of which 66 (46%) were reported as secondary in the registry (Table 5.3). The measurement variable for the secondary outcomes in the protocol was concordant with the trial registry for 54 trials (37%) and a measurement variable was not clearly specified for at least one secondary outcomes in 41 trials (28%). The analytic metric and the time point was not specified for at least one secondary outcome in 75 (82%) and 82 (69%) trials respectively and concordance could not be assessed

Table 5.3: Assessment of Discrepancies Between Trial Protocol and Trial Registry with the Individual Trial as the Unit of Analysis

	Frequency (%) [95% CI]
PRIMARY OUTCOMES (n=147)	
Classification	
Reported as Primary Outcome in Registry	132 (90) [84-94]
Not Reported as Primary Outcome in Registry	15 (10) [6-16]
Measurement Variable	
Registry Entry Concordant with Protocol	128 (87) [81-92]
Registry Entry Not Concordant with Protocol	16 (11) [7-17]
Not Specified in Protocol for at Least 1 Primary Outcome in Trial	3 (2) [0.6-6]
Analytic Metric (n=101)	
Registry Entry Concordant with Protocol	59 (58) [48-68]
Registry Entry Not Concordant with Protocol	26 (26) [18-35]
Not Specified in Protocol for at Least 1 Primary Outcome in Trial	16 (16) [10-24]
Time Point (n=115)	
Registry Entry Concordant with Protocol	90 (78) [70-85]
Registry Entry Not Concordant with Protocol	6 (5) [2-11]

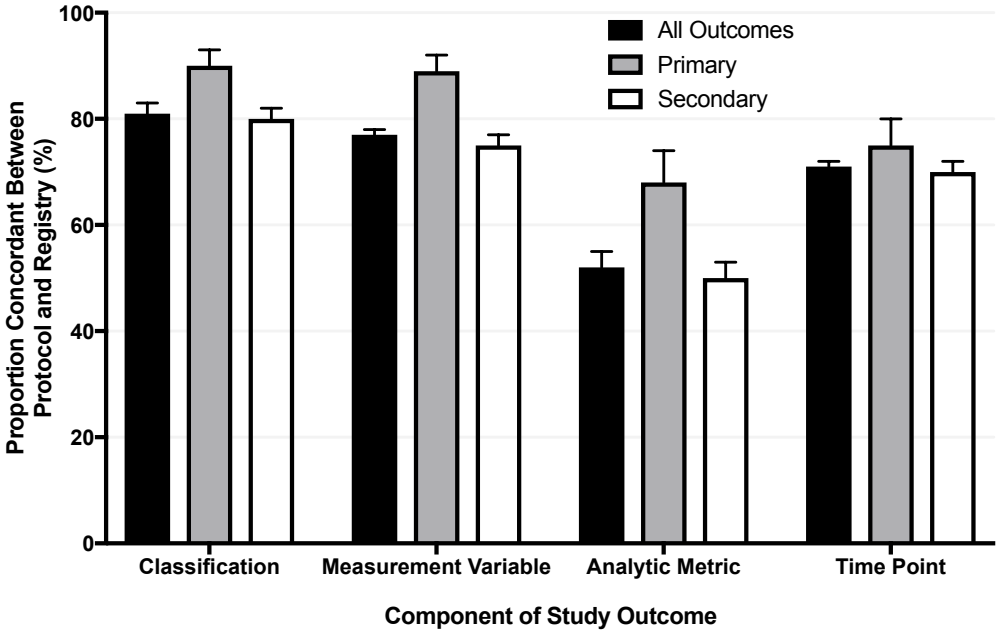
Not Specified in Protocol for at Least 1 Primary Outcome in Trial	17 (17) [11-25]
SECONDARY OUTCOMES (n=145)*	
Classification (n=144)	
Reported as Secondary Outcome in Registry	66 (46) [38-54]
Not Reported as Secondary Outcome in Registry	78 (54) [46-62]
Measurement Variable	
Registry Entry Concordant with Protocol	54 (37) [30-45]
Registry Entry Not Concordant with Protocol	50 (34) [27-43]
Not Specified in Protocol for at Least 1 Secondary Outcome in Trial	41 (28) [21-36]
Analytic Metric (n=92)	
Registry Entry Concordant with Protocol	2 (2) [0.5-8]
Registry Entry Not Concordant with Protocol	15 (16) [10-26]
Not Specified in Protocol for at Least 1 Secondary Outcome in Trial	75 (82) [72-88]
Time Point (n=119)	
Registry Entry Concordant with Protocol	27 (23) [16-31]
Registry Entry Not Concordant with Protocol	10 (8) [5-15]
Not Specified in Protocol for at Least 1 Secondary Outcome in Trial	82 (69) [60-77]

*Assessments of concordance could only be performed if each item of interest (classification, measurement variable, analytic metric and time point) was reported for all primary and secondary outcomes in the trial protocol. Any trials in which at least one outcome did not report an item of interest was classified as "Not Specified in Protocol for All Outcomes". Any studies where an analytic metric or time point did not apply for all outcomes were excluded. **The registry entry was considered to be not concordant with the protocol if the item of interest was unclear in the registry or if the item of interest or outcome was not mentioned in the registry.

Discrepancies Between Clinical Trial Protocols and Trial Registries: Analysis at the Level of the Study Outcome

Overall, 3160 study outcomes were reported in at least one of the trial protocol or trial registry and were compared between the two sources. There was a median of 37 study outcomes (IQR: 16-62) per trial. For RCTs the median number of study outcomes was 30 (16-56) and for NRTs the median number of study outcomes was 50 (17-269). Broadly, concordance between the two sources was greater for the outcome classification and measurement variable of the study outcome compared to the analytic metric and time point of the study outcome (Figure 5.2). Furthermore, agreement between the trial protocol and registry for primary outcomes was only modestly better than for secondary outcomes (Figure 5.2).

Figure 5.2: Concordance in Study Outcomes Between Trial Protocols and Trial Registries



The outcome classification was concordant[1] for 2520 (81%) of 3096 outcomes that were classified in the protocol. There were 12 (4%) primary outcomes specified in trial protocols that were not mentioned in the trial registry and there was 1 primary outcome that was newly added to the registry but not mentioned in the trial protocol (Table 5.4). Similarly, there were 487 (18%) secondary outcomes that were included in the trial protocol but not mentioned in the trial registry and there were 59 secondary outcomes that were newly added to the registry but not mentioned in the trial protocol (Table 5.4).

The measurement variable was concordant between the trial protocol and trial registry entry for 2299 of 2998 (column percentage: 77%) study outcomes where the measurement variable was reported in the protocol (Table 5.4). Among these 2998 outcomes, there were also 218 (7%) and 481 (16%) outcomes in which the measurement variable was unclear in the registry or not mentioned in the registry, respectively. With respect to the analytic metric, this component of the study outcome was concordant for 871 of 1668 outcomes (column percentage: 52%) where the analytic metric was reported in the protocol. Among these 1668 outcomes, there were also 167 (10%) and 350 (21%) outcomes in which the analytic metric was unclear in the registry or not reported in the registry, respectively. Finally, with respect to the time points for analysis, this component was concordant for 1727 of 2445 outcomes (column percentage: 71%) (Table 5.4) where the time point was reported

[1] Assessments for concordance were limited to outcomes where the component under assessment – classification, measurement variable, time point or analytic metric – was defined in the trial protocol. When the component of the the outcome was either not reported, unclear, or deemed not applicable based on the trial protocol, concordance could not be assessed and this outcome was excluded.

in the protocol. Among these 2445 outcomes, there were also 156 (6%) and 206 (8%) outcomes where the time point was unclear in the registry or not reported in the registry, respectively. Similar findings were noted when the assessment for concordance between the protocol and the registry was limited to primary and secondary outcomes. Concordance for primary outcomes was only modestly better than for secondary outcomes (Figure 5.2, Table 5.5 and 5.6).

Table 5.4: Assessment of Discrepancies Between Trial Protocol and Trial Registry with the Study Outcome as the Unit of Analysis (n=3160)

CLASSIFICATION					
	Protocol - Primary (n=324) (%) [95% CI]	**Protocol - Secondary (n=2772) (%) [95% CI]**	**Protocol - Other (n=3) (%) [95% CI]**	**Protocol - Not Mentioned (n=61) (%) [95% CI]**	**-**
Registry – Primary (n=303)	293 (90) [87-93]	9 (0.3) [0.2-0.6]	0 (0) [NA]	1 (2) [0.2-11]	**-**
Registry - Secondary (n=2307)	18 (6) [4-9]	2227 (80) [79-82]	3 (100) [NA]	59 (97) [87-99]	**-**
Registry - Other (n=51)	1 (0.3) [0.04-2]	49 (2) [1-2]	0 (0) [NA]	1 (2) [2-11]	**-**
Registry - Outcome Not Mentioned (n=499)	12 (4) [2-6]	487 (18) [16-19]	0 (0) [NA]	-	**-**
MEASUREMENT VARIABLE					
	Protocol - Reported (n=2998) (%) [95% CI]	**Protocol - Not Reported**	**Protocol - Unclear (n=101) (%) [95% CI]**	**Protocol - Not Applicable**	**Protocol - Not Mentioned (n=61) (%) [95% CI]**
Registry – Reported (n=2368)	2299 (col: 77 [75-78]) (cell: 73 [71-74])	**-**	14 (14) [8-22]	**-**	55 (90) [79-96]
Registry - Unclear (n=293)	218 (7) [6-8]	**-**	69 (68) [58-77]	**-**	6 (10) [4-21]
Registry – Outcome Not Mentioned (n=499)	481 (16) [15-17]	**-**	18 (18) [11-27]	**-**	-
ANALYTIC METRIC					
	Protocol - Reported (n=1668) (%) [95% CI]	**Protocol - Not Reported (n=113) (%) [95% CI]**	**Protocol - Unclear (n=371) (%) [95% CI]**	**Protocol - Not Applicable (n=956) (%) [95% CI]**	**Protocol - Not Mentioned (n=52) (%) [95% CI]**
Registry - Reported (n=995)	871 (col: 52 [50-55])	2 (2) [0.4-7]	93 (25) [21-30]	0 (0) [NA]	29 (56) [42-69]

	(cell: 40 [37-42])				
Registry – Not Reported (n=539)	350 (21) [19-23]	80 (71) [62-79]	101 (27) [23-32]	1 (0.1) [0.01-0.7]	7 (13) [6-26]
Registry - Unclear (n=326)	167 (10) [9-12]	12 (11) [6-18]	139 (37) [33-43]	0 (0) [NA]	8 (15) [8-28]
Registry - Not Applicable (n=942)	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	934 (98) [97-98]	8 (15) [8-28]
Registry - Outcome Not Mentioned (n=358)	280 (17) [15-19]	19 (17) [11-25]	38 (10) [8-14]	21 (2) [1-3]	-
TIME POINT					
	Protocol - Reported (n=2445) (%) [95% CI]	**Protocol - Not Reported (n=106) (%) [95% CI]**	**Protocol - Unclear (n=275) (%) [95% CI]**	**Protocol - Not Applicable (n=290) (%) [95% CI]**	**Protocol - Not Mentioned (n=44) (%) [95% CI]**
Registry - Reported (n=1856)	1727 (col: 71 [69-72]) (cell: 60 [58-62])	11(10) [6-18]	81 (30) [24-35]	0 (0) [NA]	37 (84) [70-92]
Registry – Not Reported (n=297)	206 (8) [7-10]	63 (59) [50-68]	26 (10) [7-14]	0 (0) [NA]	2 (5) [1-17]
Registry - Unclear (n=245)	156 (6) [5-7]	6 (6) [3-12]	79 (29) [24-34]	0 (0) [NA]	4 (9) [3-23]
Registry - Not Applicable (n=291)	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	277 (100) [NA]	1 (2) [0.2-15]
Registry - Outcome Not Mentioned (n=471)	356 (15) [13-16]	26 (25) [17-34]	89 (32) [27-38]	0 (0) [NA]	-

Percentages provided are column percentages unless otherwise indicated. Note that cell percentages for analytic metric and time point are calculated after excluding outcomes where either item was considered not applicable based on the study protocol.
*The specification of a study outcome inherently means that the measurement variable is either reported, unclear or not mentioned in either the protocol or trial registry. Therefore, there were no study outcomes under the heading of "Not Report" or "Not Applicable".

Table 5.5: Assessment of Discrepancies Between Trial Protocol and Trial Registry with the Study Outcome as the Unit of Analysis (Primary Outcomes Only, n=334)

CLASSIFICATION					
	Protocol - Primary (n=324) (%) [95% CI]	**Protocol - Secondary (n=9) (%) [95% CI]**	**Protocol - Other (n=0) (%) [95% CI]**	**Protocol - Not Mentioned (n=1) (%) [95% CI]**	**-**
Registry – Primary (n=303)	293 (col: 90 [87-93]) (cell: 88 [84-91])	9 (100) [NA]	0 (0) [NA]	1 (100) [NA]	**-**
Registry - Secondary (n=18)	18 (6) [4-9]	-	-	-	**-**
Registry - Other (n=1)	1 (0.3) [0.04-2]	-	-	-	**-**
Registry - Outcome Not Mentioned (n=12)	12 (4) [2-6]	-	-	-	**-**
MEASUREMENT VARIABLE					
	Protocol - Reported (n=330) (%) [95% CI]	**Protocol - Not Reported**	**Protocol - Unclear (n=3) (%) [95% CI]**	**Protocol - Not Applicable**	**Protocol - Not Mentioned (n=1) (%) [95% CI]**
Registry – Reported (n=296)	293 (col: 89 [85-92]) (cell: 88 [84-91])	**-**	2 (67) [0.3-100]	**-**	1 (100) [NA]
Registry - Unclear (n=26)	25 (8) [5-11]	**-**	1 (33) [0.07-100]	**-**	-
Registry – Outcome Not Mentioned (n=12)	12 (4) [2-6]	**-**	0 (0) [NA]	**-**	-
ANALYTIC METRIC					
	Protocol - Reported (n=216) (%) [95% CI]	**Protocol - Not Reported (n=14) (%) [95% CI]**	**Protocol - Unclear (n=30) (%) [95% CI]**	**Protocol - Not Applicable (n=73) (%) [95% CI]**	**Protocol - Not Mentioned (n=1) (%) [95% CI]**
Registry - Reported	146	0 (0)	13 (43)	0 (0)	1 (100)

(n=160)	(col: 68 [61-74]) (cell: 56 [50-62])	[NA]	[26-62]	[NA]	[NA]
Registry – Not Reported (n=57)	40 (19) [14-24]	14 (100) [NA]	3 (10) [3-28]	0 (0) [NA]	-
Registry - Unclear (n=33)	19 (9) [6-13]	0 (0) [NA]	14 (47) [29-65]	0 (0) [NA]	-
Registry - Not Applicable (n=73)	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	73 (100) [NA]	-
Registry - Outcome Not Mentioned (n=11)	11 (5) [3-9]	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	-
TIME POINT					
	Protocol - Reported (n=241) (%) [95% CI]	**Protocol - Not Reported (n=13) (%) [95% CI]**	**Protocol - Unclear (n=42) (%) [95% CI]**	**Protocol - Not Applicable (n=38) (%) [95% CI]**	**Protocol - Not Mentioned (n=0) (%) [95% CI]**
Registry - Reported (n=198)	180 (col: 75 [69-80]) (cell: 61 [55-66])	0 (0) [NA]	18 (43) [28-59]	0 (0) [NA]	0 (0) [NA]
Registry – Not Reported (n=26)	10 (4) [2-8]	11 (85) [49-97]	5 (12) [5-26]	0 (0) [NA]	-
Registry - Unclear (n=60)	46 (19) [15-25]	2 (15) [3-51]	12 (29) [17-45]	0 (0) [NA]	-
Registry - Not Applicable (n=38)	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	38 (100) [NA]	-
Registry - Outcome Not Mentioned (n=12)	5 (2) [0.8-5]	0 (0) [NA]	7 (17) [8-32]	0 (0) [NA]	-

Percentages provided are column percentages unless otherwise indicated. Note that cell percentages for analytic metric and time point are calculated after excluding outcomes where either item was considered not applicable based on the study protocol.
*The specification of a study outcome inherently means that the measurement variable is either reported, unclear or not mentioned in either the protocol or trial registry. Therefore, there were no study outcomes under the heading of "Not Report" or "Not Applicable".

Table 5.6: Assessment of Discrepancies Between Trial Protocol and Trial Registry with the Study Outcome as the Unit of Analysis (Secondary Outcomes Only, n=2852)

CLASSIFICATION					
	Protocol - Primary (n=18) (%) [95% CI]	**Protocol - Secondary (n=2772) (%) [95% CI]**	**Protocol - Other (n=3) (%) [95% CI]**	**Protocol - Not Mentioned (n=59) (%) [95% CI]**	**-**
Registry – Primary (n=9)	-	9 (0.3) [0.1-0.6]	-	-	**-**
Registry - Secondary (n=2307)	18 (100) [NA]	2227 (col: 80 [79-82]) (cell: 78 [77-80])	3 (100) [NA]	59 (100) [NA]	**-**
Registry - Other (n=49)	-	49 (2) [1-2]	-	-	**-**
Registry - Outcome Not Mentioned (n=487)	-	487 (18) [16-19]	-	-	**-**
MEASUREMENT VARIABLE					
	Protocol - Reported (n=2695) (%) [95% CI]	**Protocol – Not Reported***	**Protocol - Unclear (n=98) (%) [95% CI]**	**Protocol - Not Applicable***	**Protocol - Not Mentioned (n=59) (%) [95% CI]**
Registry – Reported (n=2096)	2031 (col: 75 [74-77]) (cell: 71 [70-73])	**-**	12 (12) [7-21]	**-**	53 (90) [79-95]
Registry - Unclear (n=269)	195 (7) [6-8]	**-**	68 (69) [59-78]	**-**	6 (10) [5-21]
Registry – Outcome Not Mentioned (n=487)	469 (17) [16-19]	**-**	18 (18) [12-27]	**-**	-
ANALYTIC METRIC					
	Protocol - Reported (n=1468) (%) [95% CI]	**Protocol - Not Reported (n=104) (%) [95% CI]**	**Protocol - Unclear (n=345) (%) [95% CI]**	**Protocol - Not Applicable (n=884) (%) [95% CI]**	**Protocol - Not Mentioned (n=51) (%) [95% CI]**
Registry - Reported	739	2 (2)	80 (23)	0 (0)	28 (55)

(n=995)	(col: 50 [48-53]) (cell: 38 [35-40])	[0.4-8]	[19-28]	[NA]	[41-68]
Registry – Not Reported (n=539)	310 (21) [19-23]	71 (68) [59-77]	99 (29) [24-34]	1 (0.1) [0.01-0.8]	7 (14) [7-27]
Registry - Unclear (n=326)	150 (10) [9-12]	12 (12) [7-19]	128 (37) [32-42]	0 (0) [NA]	8 (16) [8-29]
Registry - Not Applicable (n=942)	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	862 (98) [96-98]	8 (16) [8-29]
Registry - Outcome Not Mentioned (n=358)	269 (18) [16-20]	19 (18) [12-27]	38 (11) [8-15]	21 (2) [2-4]	-
TIME POINT					
	Protocol - Reported (n=2213) (%) [95% CI]	**Protocol - Not Reported (n=97) (%) [95% CI]**	**Protocol - Unclear (n=244) (%) [95% CI]**	**Protocol - Not Applicable (n=255) (%) [95% CI]**	**Protocol - Not Mentioned (n=43) (%) [95% CI]**
Registry - Reported (n=198)	1556 (col: 70 [68-72]) (cell: 60 [58-62])	11 (11) [6-20]	68 (28) [23-34]	0 (0) [NA]	36 (84) [69-92]
Registry – Not Reported (n=26)	196 (9) [8-10]	56 (58) [48-67]	23 (9) [6-14]	0 (0) [NA]	2 (4) [1-18]
Registry - Unclear (n=60)	110 (5) [4-6]	4 (4) [2-11]	71 (29) [24-35]	0 (0) [NA]	4 (9) [3-23]
Registry - Not Applicable (n=38)	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	255 (100) [NA]	1 (2) [0.3-16]
Registry - Outcome Not Mentioned (n=12)	351 (16) [14-17]	26 (27) [19-37]	82 (34) [28-40]	0 (0) [NA]	0 (0) [NA]

Percentages provided are column percentages unless otherwise indicated. Note that cell percentages for analytic metric and time point are calculated after excluding outcomes where either item was considered not applicable based on the study protocol.

*The specification of a study outcome inherently means that the measurement variable is either reported, unclear or not mentioned in either the protocol or trial registry. Therefore, there were no study outcomes under the heading of "Not Report" or "Not Applicable".

Multivariable models examined the concordance of information between trial protocols and trial registries for each component of the reporting of study outcomes. The analysis was at the level of the study outcome and was limited to outcomes where the item of interest was defined in the trial protocol. Otherwise, concordance between the trial protocol and trial registry could not be assessed and the outcome was excluded for the analysis for that item of interest. Pharmacologic trials were less likely to have a discrepancy between the trial protocol and trial registry for the classification of the study outcomes (aRR 0.27, 0.11-0.71; Table 5.7) and for the time point of the study outcome (aRR 0.16, 0.04-0.61, Table 5.10). Likewise, multicentre studies were significantly less likely to have a discrepancy between the trial protocol and trial registry for the analytic metric of study outcomes (aRR 0.52, 0.30-0.91; Table 5.9). Statistical significance was not achieved for the association between the outcomes of interest and study sample size and funding source. There were no clear patterns based on the relative risk estimates generated for these predictors (Table 5.7-5.10).

Table 5.7: Association Between Study Characteristics and Discrepancies in the Classification of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome)

	Number of Outcomes	Number with discrepancies (%)	Unadjusted RR (95% CI)	P-value	Adjusted RR (95% CI)	P-value
Intervention (%)						
Non-pharmacologic	490	163 (33)	ref	ref	ref	ref
Pharmacologic or Vaccine	2670	477 (18)	0.34 (0.17-0.69)	0.003	0.27 (0.11-0.71)	0.007
Number of Centres						
Single	578	144 (25)	ref	ref	ref	ref
Multiple	2582	496 (19)	0.82 (0.33-2.05)	0.67	0.79 (0.34-1.83)	0.58
Sample Size (per 1000 participants)	-	-	1.05 (1.01-1.09)	0.007	1.08 (0.99-1.16)	0.06
Funding						
Non-Industry	962	232 (24)	ref	ref	ref	ref
Solely or Part Industry	2112	402 (19)	0.58 (0.28-1.20)	0.14	1.08 (0.43-2.73)	0.86

"ref" is reference and RR is relative risk. Multilevel models were used to account for clustering at the level of the REC and the level of the individual trial. The models were additionally adjusted for whether the outcome was a primary, secondary or "other" outcome and whether the study was a pilot study.

Table 5.8: Association Between Study Characteristics and Discrepancies in the Reporting of the Measurement Variable of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome)

	Number of Outcomes	Number with discrepancies (%)	Unadjusted RR (95% CI)	P-value	Adjusted RR (95% CI)	P-value
Intervention (%)						
Non-pharmacologic	420	158 (38)	ref	ref	ref	ref
Pharmacologic or Vaccine	2578	541 (21)	0.55 (0.26-1.20)	0.13	0.56 (0.24-1.27)	0.17
Number of Centres						
Single	522	152 (29)	ref	ref	ref	ref
Multiple	2476	547 (22)	0.86 (0.30-2.42)	0.77	0.74 (0.26-2.12)	0.57
Sample Size (per 500 participants)	-	-	1.01 (0.92-1.09)	0.92	0.98 (0.83-1.12)	0.75
Funding						
Non-Industry	853	242 (28)	ref	ref	ref	ref
Solely or Part Industry	2059	377 (18)	0.67 (0.33-1.35)	0.26	0.81 (0.32-2.06)	0.66

"ref" is reference and RR is relative risk. Multilevel models were used to account for clustering at the level of the REC and the level of the individual trial. The models were additionally adjusted for whether the outcome was a primary, secondary or "other" outcome and whether the study was a pilot study.

Table 5.9: Association Between Study Characteristics and Discrepancies in the Analytic Metric of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome)

	Number of Outcomes	Number with discrepancies (%)	Unadjusted RR (95% CI)	P-value	Adjusted RR (95% CI)	P-value
Intervention (%)						
Non-pharmacologic	223	141 (63)	ref	ref	ref	ref
Pharmacologic or Vaccine	1445	656 (45)	0.73 (0.43-1.24)	0.24	0.74 (0.36-1.51)	0.40
Number of Centres						
Single	340	179 (53)	ref	ref	ref	ref
Multiple	1328	618 (47)	0.72 (0.46-1.13)	0.15	0.52 (0.30-0.91)	0.02
Sample Size (per 500 participants)	-	-	1.04 (0.90-1.18)	0.58	1.05 (0.90-1.19)	0.53
Funding						
Non-Industry	449	232 (52)	ref	ref	ref	ref
Solely or Part Industry	1206	559 (46)	0.87 (0.53-1.43)	0.59	1.14 (0.55-2.37)	0.73

"ref" is reference and RR is relative risk. Multilevel models were used to account for clustering at the level of the REC and the level of the individual trial. The models were additionally adjusted for whether the outcome was a primary, secondary or "other" outcome and whether the study was a pilot study.

Table 5.10: Association Between Study Characteristics and Discrepancies in the Time Point of Study Outcomes in Trial Protocols Compared to Trial Registries (Analysis at Level of Trial Outcome)

	Number of Outcomes	Number with discrepancies (%)	Unadjusted RR (95% CI)	P-value	Adjusted RR (95% CI)	P-value
Intervention (%)						
Non-pharmacologic	309	206 (67)	ref	ref	ref	ref
Pharmacologic or Vaccine	2136	512 (24)	0.35 (0.14-0.86)	0.022	0.16 (0.04-0.61)	0.007
Number of Centres						
Single	388	154 (37)	ref	ref	ref	ref
Multiple	2057	564 (27)	1.11 (0.37-3.28)	0.86	0.82 (0.28-2.47)	0.73
Sample Size (per 500 participants)	-	-	1.02 (0.87-1.17)	0.78	0.96 (0.81-1.11)	0.61
Funding						
Non-Industry	653	298 (46)	ref	ref	ref	ref
Solely or Part Industry	1706	414 (24)	1.00 (0.41-2.45)	0.99	2.17 (0.59-7.91)	0.24

"ref" is reference and RR is relative risk. Multilevel models were used to account for clustering at the level of the REC and the level of the individual trial. The models were additionally adjusted for whether the outcome was a primary, secondary or "other" outcome and whether the study was a pilot study.

Analysis of Amendments Submitted to Research Ethics Committee

A random sample of 30 trials was selected from the 147 trials where the study outcomes were compared between the trial protocol and trial registry. Among these 30 trials, there were 19 trials (63%) which had summary details on the amendments as well as updated versions of the trial protocol available. Also seven trials (23%) had summary information on the amendments available but updated protocols were unavailable. Nonetheless, the summary details for these seven trials were sufficient to decide whether study outcomes were included in the amendments. Finally, there were four trials (13%) where the summary information was unclear and there were no updated protocols available. For these four trials,

the summary details were insufficient to decide whether study outcomes were changed in an amendment.

Among the 26 trials with sufficient summary information on study amendments and/or updated protocols, there were three trials (11%) where study amendments accounted for the discrepancies identified between the trial protocol and trial registry. Of interest, there were also trials where amendments included new study outcomes that were not in the trial registry entry.

Reasons for Non Registration of Trials

As part of the application for research ethics approval in the UK, investigators are required to respond to questions about whether they would register their clinical trial if a trial registration number was not provided in the application. Among the 54 trials that were not registered in this study, 12 trials indicated that the investigator would register the trial after approval. However, a trial registry entry could not be found for these studies. Finally, for the remaining 42 trials, investigators indicated in the ethics application that they would not register their trial and for 29 of these 42 trials, a reason was provided for this decision.

Three independent themes emerged during the analysis of the reasons investigators chose not to register their trial. The most common was hesitancy among investigators to register trials of pilot studies or trials of behavioural interventions (Table 5.11). Second, there were misconceptions related to whether registration on a local registry operated by an investigator's institution fulfilled the requirement for trial registration. Finally, one study mentioned that registration would not be undertaken because the trial was unfunded.

Table 5.11: Reasons for Non-Registration of Clinical Trials

Pilot Study or Clinical Trial of Behavioural Intervention
"The sponsor does not wish to register the study on a public database as it is for exploratory purposes only"
"The research does not involve the use of an investigational medicinal product"
"The pilot study will not be registered. We plan to undertake a full scale study if the findings of this study are positive. At that stage we will register with a trials register"
Misconceptions About Local Databases Versus WHO Primary Registries
"The research will be registered on the relevant NHS research database and the NIHR Portfolio"
"There is no public database yet available at [our institution], but as soon as one is developed the research will be registered. In addition, individualised patient data will be stored in an anonymized form which will be available for metanalysis."
Misconceptions About Cost Related to Trial Registration
"This is a small unfunded pilot study which does not have the infrastructure to conform to ISRCTN guidelines and therefore will not be registered as a clinical trial"

5.3 CHAPTER DISCUSSION

This analysis provides an overview of trial registration in 19 RECs in the UK and a detailed comparison of information reported in clinical trial protocols and trial registries. The first main finding was that among 259 trials approved in 2012, 205 (79%) studies were registered. Second, the classification and measurement variable of study outcomes was more often concordant between the protocol and registry compared to the analytic metric and time point for the study outcome. Third, concordance between the trial protocol and registry was only modestly greater for primary outcomes compared to secondary outcomes. Fourth, pharmacologic trials and multicentre trials were less likely to have discrepancies between

the protocol and registry as compared to non-pharmacologic and single centre trials, respectively. Fifth, common reasons for non registration of trials included reluctance to register pilot studies or trials of behavioural interventions, as well as misconceptions about costs related to trial registration.

Clinical trial registration has emerged as an important strategy to improve the conduct and accountability of research. However, the utility of trial registers is contingent on ensuring that all trials are registered and that the information provided in trial registries is an accurate reflection of the apriori intentions of investigators. With respect to trial registration, 79% of trials approved by 19 RECs in the UK in 2012 were registered. This proportion was higher than in prior studies (17,18) and may reflect the use of clinical trial protocols to obtain the estimates in this thesis rather than published studies. Clinical trial protocols reflect the current research regulatory environment, and the higher estimates in this thesis may also reflect recent policy changes regarding trial registration in the UK. In particular, trial registration was recently included as a mandatory requirement for receipt of ethical approval for all clinical trials in the UK as of November 2013 (21). The high percentage of trials that are registered in the ASPIRE cohort may be higher than in other non-UK cohorts and may be a consequence of anticipation among researchers regarding the 2013 mandatory trial registration policy.

However, the value of trial registration is contingent on researchers and sponsors accurately and completely reporting information about their clinical trial in the registry (19). Our study is the first to compare the information in clinical trial protocols to clinical trial registries. Broadly, the concordance of the classification of the outcome and the measurement variable between the two sources (Concordance = 80%, 95% CI 80%-83% for

classification and 77%, 95% CI 75%-78% for measurement variable) was frequently better than the concordance of the analytic metric and the time point of the outcome (Concordance = 52%, 95% CI 50%-55% for analytic metric and 71%, 95% CI 69%-72% for time point). Furthermore, the concordance between the two sources was only modestly better for primary outcomes compared to secondary outcomes.

It is conceivable that these discrepancies reflect an oversight by investigators who quickly completed the trial registry without considering if the details were sufficient. Indeed, this was the retort by authors and journal editors whose clinical trials were assessed by the COMPare project. For instance, it was argued that discrepancies were fully explained because an unpublished study protocol was updated by the authors but this change was not reflected in the public trial registry. However, such oversights undermine the fidelity of information included in trial registries and seriously weaken the benefit that should be accrued from mandatory trial registration policies. Furthermore, a shift away from a public record about clinical trial conduct to a reliance on unpublished study protocols is regressive and reduces transparency and accountability.

Of interest, this analysis also found that pharmacologic trials and multicentre trials were less likely than their respective counterparts to have discrepancies between the trial protocol and trial registry for study outcomes. The discrepancies included changing the classification of an outcome (primary outcome to secondary outcome or vice versa) as well as changes in the analytic metric and changes in the time point. The relative risk estimates were robust and the confidence intervals were well removed from the line of no effect. It is not clear from the current study whether these discrepancies in study outcomes will be associated with statistically significant results in subsequent publications. Theoretically,

under-specification of study outcomes in the trial registry may allow for more flexibility in the analysis and interpretation of results. Although this thesis could not examine this theory, it is noteworthy that meta-epidemiological studies have demonstrated that single-centre trials show larger treatment effects for binary and continuous outcomes than multi-centre trials (22,23). While there may be several reasons for larger treatment effects in single centre trials, including the assessment of a homogenous population or the expertise of the clinical trial investigators (22,23), the extent of outcome switching in single centre studies may be a contributing factor and requires further study.

Finally, a modest proportion (20%) of trials were not registered which may be in part due to the reluctance of researchers to embrace trial registration and misconceptions about trial registries. For instance, some researchers were hesitant to register their trial because their clinical trial was designed as a pilot study or involved the use of a non-pharmacologic intervention. This reluctance among investigators likely reflects inconsistency in trial registration policies among research regulatory authorities. For instance, there is inconsistency between policies in the US and the UK as to whether all trials must be registered or only studies of pharmacologic and device interventions (16,24). However, pilot studies and trials of behavioural interventions are no different from pharmacologic trials in that these studies are an important source of clinical and public health evidence. These trials also rely on the voluntary participation of trial participants, who consent with the expectation that the trial will be conducted with rigor and their involvement will allow for the advancement of scientific knowledge (25). Accordingly, registration should be expected for all trials, irrespective of their status as a pilot study or the type of intervention. This should also be reflected in legislation related to trial registration.

Another reason some trials were not registered was concern about the costs of registration. Although some trial registries require a fee for registration of trials, the large majority of WHO primary trial registries, including clinicaltrials.gov, are free to researchers. Accordingly, there are no financial barriers to ensuring clinical trials are registered and renewed efforts are required from research regulatory authorities and ethics committees to dispel these myths.

Implications for Clinicians, Policy Makers and Future Research

The findings in this results chapter have important implications for the potential benefits that can be expected from mandatory trial registration policies. Of interest, recent studies have only found modest associations between trial registration and the direction of study results. For instance, a meta-epidemiological study found larger treatment effects in trials that were not prospectively registered (ROR 0.81, 0.65-1.02; ROR <1 indicates larger effects in trials not prospectively registered) (26). Cross-sectional studies have produced divergent results, with one study finding no association between trial registration and positive study findings in cancer trials and a second small study finding an inverse association between trial registration and positive study findings in cardiovascular trials (27,28). Whether the documented discrepancies between the protocol and trial registry are associated with positive study outcomes is not clear from this thesis and warrants further study.

Given the extent of discrepancies in study outcomes between protocols and trial registries, it is unclear how the fidelity of trial registries would be ensured. Indeed, regular comparison of outcomes between protocols and registries would be a herculean task and it is unclear whether any research regulatory authority – funding bodies, RECs, or journals – can

conceivably undertake this responsibility. Therefore, it would be pragmatic to request that researchers publish their study protocol prior to the recruitment of the first participant and ensure it is consistent with the registry entry. This would allow other investigators, patients and readers to have a clear timeline of the events in the selection of outcomes for clinical trials. Furthermore, it is important that journals are less permissive of the types of outcome switching documented in the COMPare trial (13). Given that the journals examined in the COMPare project were high impact general medical journals, it is conceivable that the problem of outcome switching may be dramatically more pervasive in journals that do not have the resources for rigorous peer review.

Finally, in order to ensure all trials are registered, it is necessary to dispel myths related to trial registration and to change researchers' perception of trial registration. This strategy will be important given inconsistency among different journals with respect to which trials must be registered. It may also be important to particularly advocate to journal editors, some of whom do not see the merit in trial registration. For instance, in a mixed methods study, trial registration policies in a random sample of 200 journals indexed in the CENTRAL database in 2012 were assessed (29). Furthermore, 13 editors and three publishers were interviewed about their views on trial registration (29). Among the 200 journals, 142 (71%) did not require trial registration or did not mention it on their journal website, 55 (28%) required registration and 3 (2%) encouraged registration but it was not mandatory (29). In the qualitative analysis of interviews with editors and publishers, trial registration was perceived as a barrier to remaining competitive with other journals (29). Furthermore, some editors and publishers did not believe that trial registration was relevant to small or exploratory studies (29). Stronger advocacy through institutions like the ICMJE

may be important for engaging journal editors in a constructive discourse about trial registration and encouraging more widespread adoption of the practice.

Strengths and Limitations

The primary limitation of this analysis relates to the availability of study amendments. Given that multiple trials were ongoing when this thesis chapter was started, complete documentation on study amendments were not available. Furthermore, the infrastructure of the NRES limited access to amended copies of trial protocols. Nonetheless, a retrospective review of a random sample of trials showed that the discrepancies between the trial protocol and trial registry were accounted for by study amendments in 11% of the 147 trials. Notably, there were new outcomes included in amendments submitted to the REC that were not updated in the trial registry. Although these new outcomes were not added to the analysis, this suggests that the results of this thesis chapter are robust and may even be conservative. The second limitation of this thesis chapter is related to the search for studies that were not registered at the time of ethics approval. Although these searches were completed in duplicate and in accordance with published recommendations (20) for searching registries, some studies that were indeed registered may not have been identified. Finally, the analysis was limited to studies that explicitly defined primary and/or secondary outcomes in the trial protocol. This criterion resulted in the exclusion of 27 studies and it is unclear if the findings are generalizable to trials that do not clearly define a primary or secondary outcome.

Conclusion

In summary, 80% of clinical trials approved by RECs in the UK in 2012 were registered. However, the concordance in the reporting of outcomes between trial protocols and trial registries was modest. Efforts to ensure all trials are registered and to ensure the fidelity of information in trial registries are needed.

6

RESULTS

THEME: SPECIAL TOPICS

Assessment of the Epidemiology and Design of Multi-Arm Trials

The prototypical RCT is a two arm parallel group trial comparing one intervention to another active intervention or placebo. This design is simple and can be readily used to evaluate a stand-alone pharmacological and non-pharmacological interventions. However, it may not be sufficient to answer contemporary clinical questions, which involve more complex interventions. Alternative clinical trial designs have become more common place, including multi-arm parallel group trials, factorial trials and split-body trials.

Among these alternative design strategies, multi-arm trials are the most commonly employed. For instance, in a methodological study of RCTs published in PubMed in December 2000 (n=383) and December 2006 (n=477), the proportion of trials that were multi-arm was 25% (97 of 383 parallel group trials) and 23% (103 of 477 parallel group

trials) in each time period respectively (1,2). In comparison, there were only 20 trials (4%) published in December 2000 and 39 trials (6%) published in December 2006 that used a factorial, split body, step wedge or "other" trial design (1).

Multi-arm trials – which are defined as parallel group trials involving three or more treatment arms – may have important benefits compared to a conventional two arm trial. First, investigators can examine more than one clinical trial question at a time using a multi-arm trial (3,4). This allows for improved research efficiency and reduced costs. Multi-arm trials in oncology also recruit faster than two-arm trials (5), perhaps due to the increased opportunity for participants to be assigned to an active intervention. Whether this is applicable to other clinical areas is unclear.

However, multi-arm trial designs are methodologically and statistically more complex than standard two-arm randomized trials (6-10). This is primarily due to the number of treatment arms under examination and the potential number of comparisons that investigators can conduct. The number of comparisons increases considerably with each additional intervention arm in a multi-arm trial. For instance, in a three-arm trial, seven potential comparisons are possible (3). The issue of multiple treatment comparisons – also known as multiplicity – must be considered in the rationale, design and analysis of multi-arm trials (11).

First and foremost, protocols for multi-arm trials should clearly identify all treatment arms in the study objectives (11). As well, comparisons between the treatment arms should be pre-specified and the primary comparison of interest should be clearly identified (11). Finally, all treatment arms should be identified in plans to stop the trial early. If the study

objective, primary comparisons or stopping rules are unclear, there is an opportunity for researchers to select the comparison that provides the most favorable result (11).

Whether and how multiplicity should be addressed at the analysis stage is more controversial. Each treatment comparison assessed in a multi-arm trial corresponds to a null-hypothesis which must be evaluated (12). For instance, consider a multi-arm trial of treatment A versus B versus Placebo. If treatment A is compared to placebo in a pairwise comparison, the null hypothesis is that there is no difference between treatment A and placebo and the alternate hypothesis is that there is a difference between treatment A and placebo. The significance level at which each null hypothesis is tested is the type I error rate or alpha, which is typically set at 5% (12). The type I error rate is the probability of incorrectly rejecting the null hypothesis, which would be a false positive result (12,13). Two forms of type I error rates should be considered: the probability of obtaining at least one false positive result, known as the family wise error rate (FWER) and the probability of obtaining multiple false positive results across all comparisons in a multi-arm trial, known as the multiple superior false positive rate (MSFP) (14). In a multi-arm trial, each pairwise comparison could be considered to be an additional opportunity for a false positive result but there is extensive debate about whether the FWER and MSFP is increased in all multi-arm trials (4,11,12,15-17).

For instance, researchers have argued that where a multi-arm trial compares two distinct interventions to a common control and the comparisons do not support the same hypothesis, the FWER would *not* be increased and the MSFP is not relevant (14). Consider a multi-arm trial comparing intervention A versus B versus C (control), where the only comparisons are between each intervention and the control group (A versus C; B versus C).

Furthermore, the success of A is a completely separate and unrelated research hypothesis from the success of B; indeed, A and B are distinct and independent interventions. In this instance, simulation studies have shown that the FWER in a multi-arm trial of two independent interventions sharing a common placebo group is 9.1% and is not increased compared to the FWER in two separate two-arm trials examining the same comparisons (9.8%) (14). Also, if the same interventions were compared to placebo using separate two-arm trials, there would be no concern about an increase in the FWER (14). The MSFP – the probability of obtaining two or more false positive results – is always increased in multi-arm trials because of the chance of having a study population that behaves differently than the true population. However, the MSFP is not relevant when interventions are distinct and act through unique independent pathways because the treatment effect of the interventions relative to the control group will be unrelated (14).

In most other designs of multi-arm trials, a type I error adjustment is needed. Consider the same trial of two interventions with a shared control group (A versus B versus C). If the interventions were compared to the control group (A versus C and B versus C) and were also compared to one another (A versus B), there are two opportunities for false positives for each intervention and therefore the FWER and MSFP are increased and relevant (14). A type I error adjustment is needed in this instance. Similarly, if comparisons in a multi-arm trial involve related interventions, the FWER is increased and MSFP is relevant (14). Consider again a multi-arm trial comparing intervention A1 versus A2 versus C, where A1 and A2 are different doses of the same intervention and are compared to intervention C (A1 versus C; A2 versus C). As each of these comparisons is related to the same primary question and the interventions involve related mechanisms of action, the MSFP will be relevant (14). Taken

together, the type I error rate in a multi-arm trial is context specific and depends on the primary objective of the trial, the number of comparisons needed to answer the primary research question and the interventions examined within the trial.

There are a range of strategies researchers can employ to address an increased FWER and MSFP in a multi-arm trial. First, if the primary research question is to examine the overall difference between the interventions in the trial as opposed to any specific pairwise comparison, a global test that simultaneously compares all groups can be used. For instance, in the three-arm trial comparing A versus B versus C, the global test would assess the null hypothesis of whether A=B=C (11). This approach circumvents the issue of multiplicity. Similarly, in a multi-arm trial of increasing doses of the same intervention, the issue of multiplicity could also be avoided by testing for a dose response relationship.

Second, researchers could use a prioritized sequence of tests to conduct their analysis. In this approach, a hierarchical list of the planned comparisons is pre-specified in the trial protocol (11,18). Each comparison would be assessed using the study level type I error rate, typically set at 5% (11,18). However, researchers would only assess a comparison if the null hypothesis for the analysis directly preceding it was rejected (11,18). This limits the number of null hypotheses that are tested and prioritizes the comparisons that of greatest interest.

Finally, researchers can adjust the type I error rate in their study based on the number of comparisons of interest. The most common approach for this is through a Bonferroni adjustment, where the type I error is divided by the number of comparisons. The quotient is then used as the new type I error rate (15). For instance, consider a three arm trial comparing treatment A1 versus A2 versus C. If there were two comparisons of interest with a study level type I error rate of 0.05, then the type I error rate used for each of the three comparisons

would be 0.025 or 0.05/2 (15). This approach is more conservative than the aforementioned methods and to compensate, researchers may increase the sample size in their trial (15).

Finally, it is notable that multiplicity and concerns about the type I error rate can occur in other situations outside of a multi-arm trial (11). For instance, the assessment of multiple primary and secondary outcomes and the use of multiple sub-group comparisons can raise multiplicity concerns. However, these issues are beyond the scope of this thesis chapter.

Considering the potential trade-offs of multi-arm trials, there is surprisingly limited information on their conduct and reporting, perhaps because of the lack of consensus on the design, conduct and analysis of these types of trials. To date, only one methodological study has examined a cohort of published multi-arm trials. The authors noted important deficiencies in reporting, including a lack of reporting of baseline characteristics for each intervention arm and a lack of reporting of pre-specified global comparison tests and pairwise comparisons (3). However, this study examined RCTs published in 2009, and so, may not be wholly representative of the conduct of contemporary RCTs. Analyses of other trial documents, such as trial protocols, are necessary to obtain more insight into the design, analysis and interpretation of multi-arm trials. The objectives of this thesis chapter were four-fold. First, it was to ascertain the epidemiology of multi-arm trials and determine how frequently multi-arm trials are conducted. Second, it was to to characterize the planned design and analysis of multi-arm trials in the ASPIRE, Swiss and ESORT cohort (see Chapter 2, Methods, for cohort details). Third, it was to assess the frequency of type I error rate adjustment in multi-arm trials in the ASPIRE, Swiss and ESORT cohort compared to what would be recommended based on a decision tool of simulation studies of multi-arm trials.

Finally, it was to assess whether there was an association between funding source (industry versus non-industry funding) and the reporting of the planned analysis of the multi-arm trial.

6.1 CHAPTER SPECIFIC METHODS

Systematic Search to Determine the Epidemiology of Multi-Arm Trials Based on Published Cohort Studies of Clinical Trial Protocols

A systematic search of PubMed (inception to June 2017) was conducted to identify cohort studies of clinical trial protocols. The outcome of interest was either of the following details about the epidemiology of the cohort: 1) the number of RCTs with three or more treatment arms; 2) the number of parallel group RCTs with three or more arms; 3) or the number of parallel group RCTs with three or more arms, excluding pilot studies (the definition used in this chapter for multi-arm trials). The title and abstract of all records retrieved from the PubMed search were reviewed for eligible studies. The search strategy is provided below:

Search Strategy: (randomized controlled trials as topic[mesh] OR clinical trials as topic[mesh] OR randomised[tiab] OR randomized[tiab]) AND (protocols[tiab]) AND (cohort studies[mesh])

Analysis of the ASPIRE, Swiss and ESORT Cohort

Definitions

Multi-arm Trials: Randomized parallel group trials involving three or more treatment arms. An example of a multi-arm trial is a clinical trial comparing the treatment of ischemic stroke in adults randomized in a 1:1:1 ratio to aspirin, warfarin or a new medication. Two-arm parallel group trials, factorial trials, crossover trials and split body trials were excluded. Pilot studies were also excluded from this study because the goal of these studies was on feasibility and recruitment and the study groups are often not formally compared. Excluding pilot studies ensured that this chapter was not underestimating the frequency with which appropriate analytical strategies were employed to address the statistical issues in multi-arm trials.

Data Extraction

Information on the general characteristics of multi-arm trials was extracted, as described in the Methods Chapter. In addition, the following methodological and statistical details were assessed and extracted from each study. First, the trial structure was determined based on the type of intervention employed in the trial. Broadly, there are three ways in which interventions are incorporated in multi-arm trials: 1) multiple active interventions; 2) combinations of active interventions; 3) different doses of the same intervention. The structure of each multi-arm trials is important to understand the potential number of comparisons and the need for adjustment of the type I error rate.

The allocation ratio to treatment arms in each trial was noted as well as if a reason was provided for an unequal allocation ratio. The number of planned primary comparisons were determined and the protocol was assessed for whether all study arms were mentioned in the study objectives and in the planned analysis (Primary or Secondary). The sample size

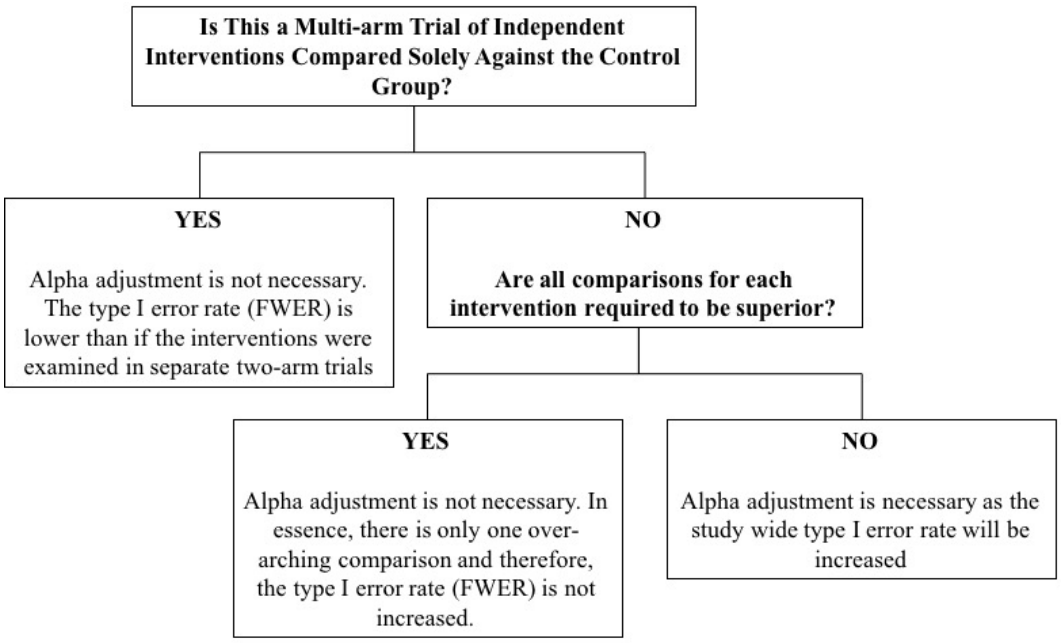
calculation was assessed to determine if the planned primary comparison was used in the calculation.

Each trial protocol was assessed to determine whether multiple testing would be performed and if yes, whether a global test would be performed and whether there was any planned adjustment for the type I error rate or the sample size calculation.

Decision Tool to Determine the Need for Adjustment for Multiple Testing

We employed a published decision tool to assess whether multi-arm trials should perform an alpha adjustment to control the type I error rate. The decision tool is based on simulation studies comparing the FWER and MSFP in multi-arm trials versus if the interventions were studied in separate two-arm trials. The tool involves four decision points (Figure 6.1).

Figure 6.1: Decision Tool to Determine the Need for Type I Error Adjustment in Multi-Arm Trials



Decision Point 1: Is This a Multi-Arm Trial of Independent Interventions Compared Solely Against the Control Group?

Where a multi-arm trial involves unrelated interventions and multiple comparisons do not provide accumulating support for any single intervention, the probability of making a single type I error rate is less than if the comparisons were conducted in separate two arm trials. Therefore, an alpha adjustment to control the FWER is not needed. If this is not the case, decision point 2 (below) should be considered.

Decision Point 2: Are All Comparisons Required to Show Superiority in Favor of the Intervention

In multi-arm trials with more than one comparison for a single intervention, the FWER and MSFP are increased. Therefore, where the comparisons in multi-arm trials provide accumulating evidence supporting an intervention, it is important to consider whether all comparisons must show superiority. If it is NOT necessary for all comparisons to show superiority, then an alpha adjustment to control the FWER and MSFP is needed. If it is necessary for all comparisons to show superiority, then in effect, there is only one overarching comparison being tested. Therefore, alpha adjustment to control the FWER and MSFP is not needed.

Using the decision tool, we identified which RCTs warranted an alpha adjustment to control the type I error rate. We compared this to the number of RCTs that reported a plan to adjust their alpha in their protocol or in the published article.

Data Analysis

A descriptive analysis was performed, as detailed in the Methods chapter. The primary focus of this analysis was on the combined ASPIRE and Swiss clinical trial protocol cohort and the analysis of the ESORT data was considered as a secondary supportive analysis and interpreted within the context of the results of the ASPIRE and Swiss cohort. This is because reliance on the published report of RCTs may give biased results about the conduct of the trial due to incomplete reporting.

Poisson regression was used to assess the association between general study characteristics – namely industry funding – and the reporting for the following items related

to the analysis of multi-arm trials: whether a global test was performed or explicitly stated as no global test performed; whether adjustment for the type I error rate was appropriately performed or addressed for trials where an adjustment was needed based on the decision tool; and whether the sample size was adjusted for multiplicity. These outcomes were of interest as they have direct implications for the results of multi-arm trials and the potential for bias. For the outcome of whether the sample size was adjusted for multiplicity, regression models were unstable due to the small number of trials as well as small number of events. Therefore, results were not reported for this outcome. Finally, the regression model for the combined ASPIRE and Swiss cohort accounted for clustering at the level of the country (UK and Switzerland). The only predictor of interest included in multivariable models were funding source (industry versus non-industry) and models were otherwise unadjusted.

The results of the regression model for the combined ASPIRE and Swiss cohort and for the ESORT cohort were displayed separately and pooled using a random effects meta-analysis. Heterogeneity was quantified as the I^2 statistic.

6.2 CHAPTER RESULTS

Prevalence of Multi-Arm Trials

The ASPIRE cohort included 21 (11%) multi-arm trials (Figure 6.2) out of 259 clinical trials and 193 RCTs. The Swiss cohort included 46 (17%) multi-arm trials, out of 273 RCTs but data extraction was only available for 32 of the 46 trials. Finally, the ESORT cohort included 133 (12%) published multi-arm trials, out of 1122 RCTs. With respect to the search for cohort studies of clinical trial protocols, 1816 records were obtained and a screen

of titles and abstracts was performed (Figure 6.3). This process identified 40 articles for full text review and 3 studies eligible for inclusion.

The published cohort studies of clinical trial protocols included trials approved from 1994 to 2003 (Table 6.1). All cohorts involved protocols approved by European RECs. The prevalence of RCTs with three of more intervention arms ranged from 17% to 25% in the published cohort studies; the corresponding percentage in the ASPIRE, Swiss and ESORT cohorts were 23%, 25% and 22% respectively. The prevalence of parallel group RCTs with three or more intervention arms (including pilot studies) was reported in only one published study of clinical trial protocols and was 22%, which was slightly higher than the ASPIRE, Swiss and ESORT cohorts. No published cohorts of clinical trial protocols specifically focused on multi-arm trials as defined in our current study: parallel group RCTs with three or more intervention arms, excluding pilot studies.

Figure 6.2: Systematic Search for Cohort Studies of Clinical Trial Protocols

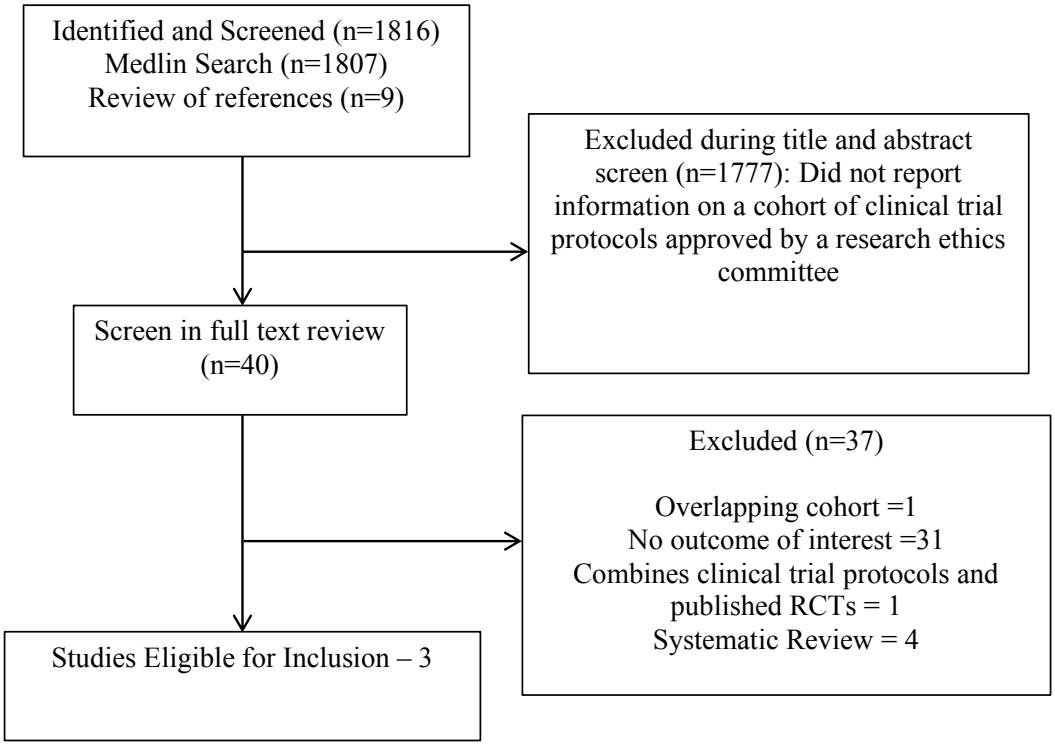


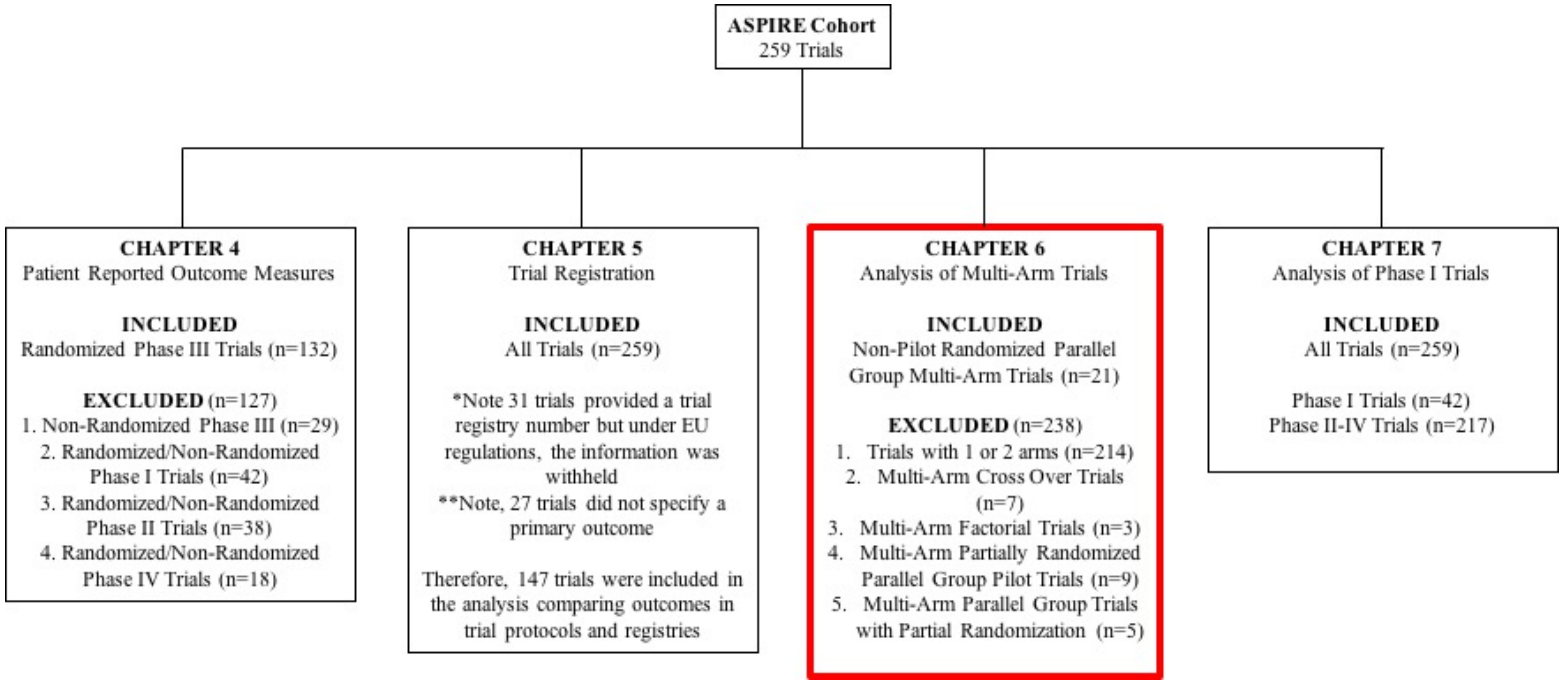
Table 6.1: Epidemiology of Multi-Arm Trials in the ASPIRE Cohort, Swiss Cohort, Published Studies of Clinical Trial Protocols and the ESORT Cohort

Trial Characteristic	Unpublished Studies of Clinical Trial Protocols		Published Studies of Clinical Trial Protocols			Cross-Sectional Study of Published RCTs
	ASPIRE Cohort	Swiss Trial Protocol Cohort	Chan et al.[1]	Blumle et al.	Bonfill et al.	ESORT Cohort (Published December 2012)
Date of Trial Protocol Approval	2012	2012	1994-1995	2000	1999-2003	N/A
Location of Ethics Committee	United Kingdom	Switzerland	Denmark	Germany	Spain	N/A
Cohort Type	All approved clinical trials by RECs	All approved clinical trials by RECs	All approved clinical trials by RECs	All approved clinical trials by RECs	Clinical Trials for Cancer Medications	RCTs published and indexed in PubMed in December 2012
Total Number of RCT Trial Protocols or Published RCTs	193	273	102	52	303	1112
Number of RCTs with 3 or more intervention arms (%)	45 (23)	67 (25)	Not Reported	13 (25)	51 (17)	220 (22)
Number of parallel group RCTs with 3 or more intervention arms (%)	35 (18)	46 (17)	23 (22)	Not Reported	Not Reported	174 (16)
Number of parallel group RCTs with 3 or more intervention arms (excluding pilot studies, %)**	21 (11)	46 (17)	Not Reported	Not Reported	Not Reported	133 (12)

**Definition used to identify multi-arm trials in this thesis

1. Chan A-W, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ 2008;337(dec04 1):a2299–a2299. doi:10.1136/bmj.a2299.
2. Blümle A, Meerpohl JJ, Rücker G, Antes G, Schumacher M, Elm von E. Reporting of eligibility criteria of randomised trials: cohort study comparing trial protocols with subsequent articles. BMJ 2011;342(apr05 2):d1828–d1828. doi:10.1136/bmj.d1828.
3. Bonfill X, Ballesteros M, Gich I, Serrano MA, García López F, Urrutia G. Description of the protocols for randomized controlled trials on cancer drugs conducted in Spain (1999-2003). PLoS ONE 2013;8(11):e79684. doi:10.1371/journal.pone.0079684.

Figure 6.3: Identification of Studies from the ASPIRE Cohort That Were Included in the Analysis of Multi-Arm Trials



General Characteristics of Included Trials

The general characteristics of included trials from each cohort are detailed in Table 6.2. Specific to the ASPIRE cohort, 14 of 21 trials (67%) involved a pharmacologic intervention, 20 (95%) were multicentre, and 15 (71%) had three intervention arms. The median sample size was 420 (IQR 200-1000) and the majority were industry funded (16 of 21 trials, 76%). The characteristics of the ASPIRE and Swiss cohort were similar except that a larger percentage of the trials in the Swiss cohort involved pharmacologic interventions (84% for the Swiss cohort and 66% for the ASPIRE cohort). Finally, the characteristics of the ESORT cohort and the ASPIRE cohort were similar except that 57 of 133 of the ESORT cohort (44%) had an unclear number of centres based on the published article and 37 trials (28%) had an unclear source of funding.

Table 6.2: General Characteristics of Multi-Arm Trials

	Clinical Trial Protocols			**Published RCTs**
	ASPIRE and Swiss Cohort Combined n (%) (n=53)	**ASPIRE Cohort n (%) [95% CI] (n=21)**	**Swiss Cohort n (%) [95% CI] (n=32)**	**ESORT Study n (%) [95% CI] (n=133)**
Intervention				
Drug	41 (77)	14 (66)	27 (84)	89 (67)
Device	2 (4)	1 (5)	1 (3)	8 (6)
Procedure/Surgery	6 (11)	4 (18)	2 (6)	14 (11)
Behavioural/Lifestyle/Education	3 (6)	2 (10)	1 (3)	22 (17)
Other	1 (2)	0 (0)	1 (3)	0 (0)
Study Centres				
Single	4 (8)	1 (5)	3 (10)	43 (33)
Multiple	49 (92)	20 (95)	29 (90)	29 (22)
Unclear	0 (0)	0 (0)	0 (0)	57 (44)
Number of Arms				
3	32 (60)	15 (71)	17 (53)	97 (73)
4	11 (21)	2 (10)	9 (28)	24 (18)

>4	10 (19)	4 (19)	6 (19)	12 (9)
Sample size				
Median (Inter-quartile range)	200 (368-600)	420 (200-1000)	203 (311-600)	84 (42-194)
Funding				
Non-Industry Funded	9 (17)	5 (29)	4 (12)	53 (40)
Industry Funded	44 (83)	16 (76)	28 (88)	43 (32)
Unclear	0 (0)	0 (0)	0 (0)	37 (28)

Methodological Items Related to the Design of Multi-Arm Trials

For the combined ASPIRE and Swiss Cohort, the most common structure for multi-arm trials was the comparison of multiple doses or intensities of the same intervention (Table 6.3). Furthermore, 36 of 53 trial protocols (68%, 95% CI 54%-79%) mentioned all study arms in the study objectives, 40 of 53 trial protocols (75%, 95% CI 62%-85%) mentioned all study arms in the planned primary analysis and 48 of 53 trials protocols (91%, 95% CI 79%-96%) mentioned the planned primary comparisons in the sample size calculation. Similar findings were noted for trials in the ESORT analysis. However, a smaller percentage in ESORT reported the planned primary comparison in the sample size calculation (39% for ESORT versus 91% in the ASPIRE and Swiss cohorts combined) and reported a sample size calculation (63% in ESORT versus 100% in the ASPIRE and Swiss cohorts combined).

Table 6.3: Methodological Items Related to the Design of Multi-Arm Trials

	Clinical Trial Protocols			Published RCTs
	ASPIRE and Swiss Cohort Combined n (%) [95% CI] (n=53)	**ASPIRE Cohort n (%) [95% CI] (n=21)**	**Swiss Cohort n (%) [95% CI] (n=32)**	**ESORT Study n (%) [95% CI] (n=133)**
Allocation Ratio				
Equal	39 (74) [60-84]	16 (76) [54-90]	23 (72) [54-85]	125 (77) [88-97]
Unequal, reason provided	5 (9) [4-21]	2 (10) [2-31]	3 (9) [3-25]	1 (1) [0-5]
Unequal, no reason provided	6 (11) [5-23]	0 (0) [NA]	6 (19) [9-36]	5 (3) [2-9]
Not stated	3 (6) [2-16]	3 (14) [5-36]	0 (0) [NA]	32 (19) [18-32]
Structure				
Multiple Independent Active Interventions Versus a Common Control	14 (26) [16-40]	9 (42) [24-64]	5 (16) [7-32]	65 (49) [40-57]
Multiple Doses or Intensities of the Same Intervention	33 (62) [49-74]	8 (38) [20-60]	25 (78) [61-89]	52 (39) [31-48]
Combination of Active Treatments	6 (11) [5-23]	4 (20) [7-41]	2 (6) [2-22]	16 (12) [8-19]
All Arms Mentioned in Study Objectives				
Yes	36 (68) [54-79]	15 (71) [49-87]	21 (66) [48-80]	131 (98) [94-100]
No	17 (32) [21-46]	6 (29) [13-51]	11 (34) [20-52]	2 (2) [0-6]
All Arms Mentioned in Planned Analyses				
Yes	40 (75) [62-85]	16 (76) [54-90]	24 (75) [57-87]	98 (74) [66-80]
No	13 (25) [15-38]	5 (24) [10-46]	8 (25) [13-43]	8 (6) [3-12]
Primary Analysis Not Stated	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	27 (20) [14-28]
Planned Primary Comparison Mentioned in Sample Size Calculation				
Yes	48 (91)	18 (86)	30 (94)	52 (39)

	[79-96]	[64-95]	[78-98]	[31-48]
No, Planned Primary Comparison Not Mentioned in Sample Size Calculation	5 (9) [4-21]	3 (14) [5-36]	2 (6) [2-22]	32 (24) [18-32]
No Sample Size Calculation Reported	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	49 (37) [29-45]

Methodological Items Related to Multiple Testing in Multi-Arm Trials

Multiple testing was planned in 43 of 53 trial protocols in the combined ASPIRE and Swiss cohort (Table 6.4). Among the 43 trial protocols which included multiple testing in their statistical analysis plan, 10 (23%, 95% CI 13%-38%) intended to perform a global comparison test. With respect to the type I error rate, 16 of 43 trial protocols (37%, 95% CI 24%-52%) intended to perform an adjustment of the type I error rate to account for multiple testing and 8 (19% 95% CI 10%-33%) used an alternative strategy to control the type I error rate. Finally, 1 of 43 trial protocols (2% 95% CI 0%-15%) did not adjust the type I error rate but provided a reason, 8 trial protocols (19%, 95% CI 10%-33%) did not adjust the type and did not provide a reason and 10 trial protocols (23%, 95% CI 13%-38%) did not state whether a type I error rate would be controlled. Only 6 of 15 trials protocols (40%, 95% CI 19%-65%) which adjusted the type I error rate also adjusted their sample size calculation to maintain statistical power. Similar findings were noted based on the published RCTs in the ESORT analysis. However, a global comparison test was more often used (62% in ESORT versus 23% in ASPIRE and Swiss cohorts combined). As well, only 5% of published RCTs in the ESORT analysis reported adjusting their sample size calculation.

Table 6.4: Methodological Items Related to Multiple Testing in Multi-Arm Trials

	Clinical Trial Protocols			Published RCTs
	ASPIRE and Swiss Cohort Combined n (%) [95% CI]	**ASPIRE Cohort n (%) [95% CI]**	**Swiss Cohort n (%) [95% CI]**	**ESORT Study n (%) [95% CI]**
Multiple Testing Present	**(n=53)**	**(n=21)**	**(n=32)**	**(n=133)**
Yes	43 (81) [63-88]	18 (86) [64-95]	25 (78) [47-86]	117 (88) [81-92]
Unclear if Multiple Testing	2 (4) [1-18]	0 (0) [NA]	2 (6) [3-32]	10 (8) [4-13]
No, Dose Response Analysis	2 (4) [0-15]	0 (0) [NA]	2 (6) [1-28]	1 (1) [0-5]
No, Multiple Arms of Intervention Collapsed to Form 2-arm Trial	6 (11) [7-29]	3 (14) [5-36]	3 (9) [5-38]	4 (3) [1-8]
No, Only Two Groups Compared and Other Groups are Descriptive	0 (0) [NA]	0 (0) [NA]	0 (0) [NA]	1 (1) [0-5]
If Multiple Testing, Is a Global Comparison Test Intended?	**(n=43)**	**(n=18)**	**(n=25)**	**(n=116)**
Yes	10 (23) [13-38]	4 (22) [9-46]	6 (24) [11-44]	72 (62) [53-70]
Explicitly Stated as No Global Test to be Performed	10 (23) [13-38]	0 (0) [NA]	10 (40) [23-60]	0 (0) [NA]
Not Stated in Protocol or Published RCT	23 (53) [39-68]	14 (78) [54-91]	9 (36) [20-56]	44 (38) [30-47]
If Multiple Testing, Any Planned Adjustment for Type I Error Rate	**(n=43)**	**(n=18)**	**(n=25)**	**(n=116)**
Yes	16 (37) [24-52]	4 (22) [9-46]	12 (48) [30-67]	42 (36) [28-45]
No Type I Error Adjustment but Alternative Strategy Used (e.g. Prioritized Sequence of Tests)	8 (19) [10-33]	2 (11) [3-35]	6 (24) [11-44]	14 (12) [7-19]
No type I error Adjustment Planned/Performed AND Reason Provided	1 (2) [0-15]	0 (0) [NA]	1 (4) [1-24]	3 (3) [1-8]
No type I Error Adjustment Planned/Performed but Reason ***NOT*** Provided	8 (19) [10-33]	2 (11) [3-35]	6 (24) [11-44]	57 (49) [40-58]

TRIAL PROTOCOLS: Not Stated How and If Type I Error Rate Would be Controlled	10 (23) [13-38]	10 (55) [33-76]	0 (0) [NA]	-
If Multiple Testing and Type I Error Rate Adjusted, Was the Sample Size Calculation Also Adjusted To Maintain Statistical Power	**(n=15)**	**(n=4)**	**(n=11)**	**(n=42)**
Yes	6 (40) [19-65]	3 (75) [24-97]	5 (45) [20-73]	2 (5) [0-7]
No Sample Size Adjustment Performed*	9 (60) [35-81]	1 (25) [3-76]	6 (55) [27-80]	40 (95) [26-44]

*Published randomized controlled trials that did not perform a sample size calculation were included in this category

Comparison Between Trial Protocols and the Decision Tool for Adjustment of Type I Error Rate

Based on the decision tool, 32 of 43 multi-arm trial protocols (74%) in the ASPIRE and Swiss combined cohort required an adjustment for the type I error rate. Among these trials, 6 (19%, 95% CI 7%-32%) did not state how and if the type I error rate would be controlled and 6 (19%, 95% CI 7%-32%) stated that no adjustment would be made but did not provide a reason (Table 6.5). Similar findings were noted in the analysis of the ESORT data.

There were also 11 of 43 (26%) multi-arm trial protocols in the ASPIRE and Swiss combined cohort that did not require an adjustment for the type I error rate based on the decision tool. Only 1 (9%) of these trials intended to perform an adjustment (Table 6.5). Based on the analysis of the ESORT data, 5 of 12 RCTs (42%) that did not require an

adjustment for the type I error rate nonetheless performed an adjustment or used an alternative strategy to control the type I error rate.

Table 6.5: Comparisons Between Analysis Strategy in Trial Protocols and the Decision Tool for the Adjustment of Type I Error Rate in Multi-Arm Trials with Multiple Testing Planned

	Yes, Type I Error Adjustment Recommended n (%) [95% CI]	**No Type I Error Adjustment Recommended n (%) [95% CI]**
ASPIRE and Swiss Cohort Combined (n=43 protocols with multiple testing)	**(n=32)**	**(n=11)**
Yes	15 (47) [31-64]	1 (9) [1-44]
No Type I Error Adjustment but Alternative Strategy Used (e.g. Prioritized Sequence of Tests)	4 (13) [5-29]	4 (36) [14-66]
No type I error Adjustment Planned/Performed AND Reason Provided	1 (3) [0-19]	0 (0) [NA]
No type I Error Adjustment Planned/Performed but Reason ***NOT*** Provided	6 (19) [7-32]	2 (18) [5-51]
TRIAL PROTOCOLS: Not Stated How and If Type I Error Rate Would be Controlled	6 (19) [9-36]	4 (36) [14-66]
ASPIRE Cohort (n=18 protocols with multiple testing)	**(n=13)**	**(n=5)**
Yes	3 (23)	1 (20)
No Type I Error Adjustment but Alternative Strategy Used (e.g. Prioritized Sequence of Tests)	2 (15)	0 (0) [NA]
No type I error Adjustment Planned/Performed AND Reason Provided	0 (0) [NA]	0 (0) [NA]
No type I Error Adjustment Planned/Performed but Reason ***NOT*** Provided	2 (15)	0 (0) [NA]
TRIAL PROTOCOLS: Not Stated How and If Type I Error Rate Would be Controlled	6 (46)	4 (80)
Swiss Cohort (n=25 protocols with multiple testing)	**(n=19)**	**(n=6)**
Yes	12 (63) [40-81]	0 (0) [NA]

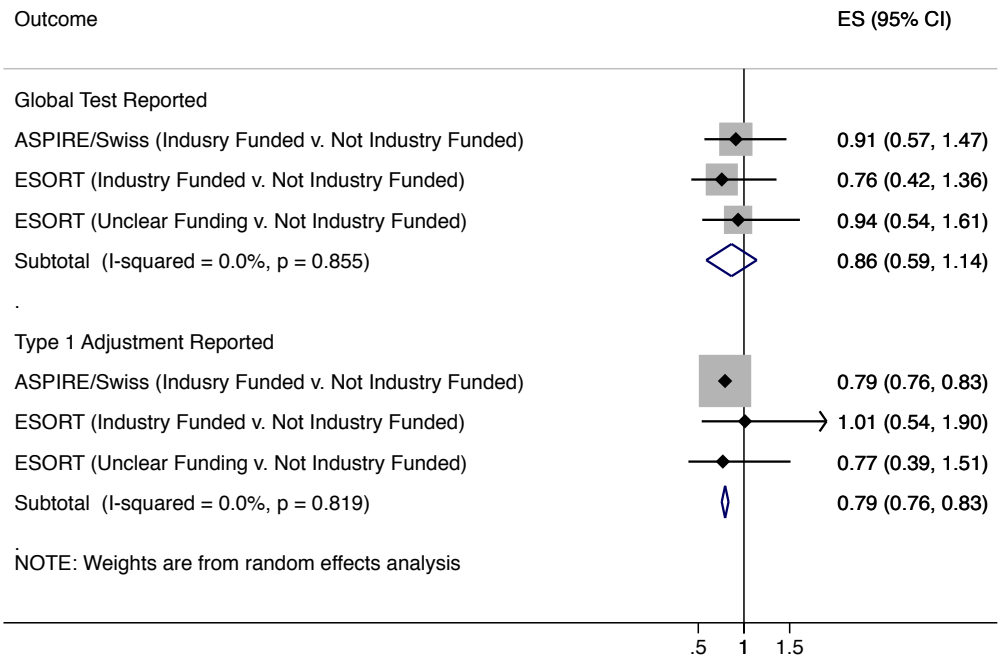
No Type I Error Adjustment but Alternative Strategy Used (e.g. Prioritized Sequence of Tests)	2 (11) [3-34]	4 (66) [27-92]
No type I error Adjustment Planned/Performed AND Reason Provided	1 (5) [1-29]	0 (0) [NA]
No type I Error Adjustment Planned/Performed but Reason ***NOT*** Provided	4 (21) [8-45]	2 (33) [8-73]
TRIAL PROTOCOLS: Not Stated How and If Type I Error Rate Would be Controlled	0 (0) [NA]	0 (0) [NA]
ESORT (n=116 published RCTs with multiple testing)	**(n=104)**	**(n=12)**
Yes	38 (37) [28-46]	4 (33) [13-62]
No Type I Error Adjustment but Alternative Strategy Used to Control Type I Error Rate (e.g. Prioritized Sequence of Tests)	13 (13) [7-20]	1 (8) [1-41]
No type I error Adjustment Planned/Performed AND Reason Provided	2 (2) [1-7]	1 (8) [1-41]
No type I Error Adjustment Planned/Performed but Reason ***NOT*** Provided	51 (49) [40-59]	6 (50) [24-76]
TRIAL PROTOCOLS: Not Stated How and If Type I Error Rate Would be Controlled	-	-

Association Between Funding Source and Reporting of Analysis Strategies in Multi-Arm Trials

In the combined ASPIRE and Swiss cohort, 17 of 36 industry funded trials (47%) and 3 of 7 non-industry funded trials (43%) that required multiple testing reported details about whether a global comparison test was intended. There was no association between industry funding and the reporting of whether a global test would be performed (RR 0.91, 0.57-1.47, Figure 6.4). With respect to performing a type I error adjustment when it was deemed necessary by the decision tool, 23 of 29 industry funded trials (79%) and 3 of 3 non-industry funded trials (100%) reported details about a type I error adjustment. Therefore, industry funded multi-arm trials were less likely than non-industry funded trials to report details about a type 1 error adjustment when it was deemed necessary based on the decision

tool (RR 0.79, 0.76-0.83, Figure 6.4). These results were consistent with the analysis of the ESORT study with no heterogeneity among study estimates (I^2=0%, Figure 6.4). However, the overall study estimate for the reporting of a type I error adjustment was primarily driven by the combined ASPIRE and Swiss cohort result.

Figure 6.4: Association Between Funding Source and Reporting of Analytical Strategies for Multi-Arm Trials



ES is effect size and is the risk ratio generated from poisson regression.

6.3 CHAPTER DISCUSSION

This chapter provides an overview of the design and analysis of multi-arm trials approved by 19 research ethics committees in the UK, 4 of 9 research ethics committee in Switzerland and 133 published multi-arm trials. The key findings were four-fold. First, the prevalence of multi-arm trials was similar across cohort studies of clinical trial protocols. Second, most multi-arm trial protocols in the combined ASPIRE and Swiss cohort accounted for all study arms in their objectives, sample size calculation and planned analysis. However, when a type I error adjustment was needed based on the decision tool, 19% of trial protocols did not state whether the type I error rate would be controlled and 19% did not provide a reason for their decision to maintain the type I error rate at 5%. Third, industry funded multi-arm trials were less likely than non-industry funded multi-arm trials to report details of type I error adjustment when it was needed based on the decision tool. Finally, the results of the analysis of published RCTs in the ESORT cohort was similar to the ASPIRE and Swiss cohort except that reporting of sample size calculations was infrequent and adjustment for multiplicity was often performed when it was not needed. This chapter therefore provides new insight into the epidemiology, design and analysis of multi-arm trials.

An important finding in this chapter relates to the prevalence of multi-arm trials. Although the percentage of multi-arm trials in the ASPIRE cohort was slightly lower than the Swiss cohort (11% versus 17%), the prevalence is still in keeping with other published studies of clinical trial protocols as well as the ESORT cohort of published RCTs (12% were multi-arm trials). In a cross-sectional analysis of published RCTs that were indexed in PubMed in December 2000 and 2006, the prevalence of multi-arm trials was 25% in 2000

(97 of 383) and 22% (103 of 477) in 2006 (1,2). Accordingly, the markedly lower prevalence of multi-arm trials in the ASPIRE and Swiss cohort and the ESORT study suggests that multi-arm trials may make up a smaller share of RCTs that are conducted or published, likely because of a proliferation of traditional two-arm parallel group trials. The less frequent use of multi-arm trials may reflect uncertainty about the strengths and limitations of this trial design. Indeed, the results of this thesis chapter noted some important gaps in the planning and analysis of multi-arm trials.

For instance, while most trials protocols mentioned all intervention arms in the study objectives and planned analysis, important methodological items related to multiple testing were inconsistently reported. For instance, in the combined ASPIRE and Swiss cohort, 32 of 43 trials required adjustment of the type I error rate based on a published decision tool (14). However, 19% of these trials protocols did not state whether the type I error rate would be controlled and 19% did not provide a reason for their decision to maintain the type I error rate at 5%. As a result, these trials are more likely to incorrectly identify a novel intervention as efficacious, potentially leading to its inappropriate use for the provision of clinical care. The reasons for omitted discussion about the type I error rate in multi-arm trial protocols is unclear. It may be that researchers are unfamiliar with methodological issues that are relevant to multi-arm trials or were unsure about whether an adjustment for multiplicity was warranted in their trials. Indeed, the existing statistical literature on the utility of adjustment of the type I error rate is often inconsistent (4,12,15-17).

Notably, in the combined ASPIRE and Swiss cohort, there were 15 trials which intended to adjust or control the type I error rate in their analysis. However, only 6 of these trials (40%, 95% CI 19%-65%) also adjusted their sample size calculation. Without an

adjustment of the sample size to reflect a new type I error rate, multi-arm trials may be underpowered and may be more prone to type II error: incorrectly failing to reject the null hypothesis (14). Type II errors have the effect of incorrectly concluding that there is no benefit from a new intervention and may therefore reduce clinical treatment options available to patients. The reduction in statistical power will depend on the number of comparisons in the multi-arm trial and the degree of adjustment of the type I error rate (14). If the increase in sample size is prohibitive, the trial may be better designed as a traditional two-arm trial or using alternative designs. Accordingly, there is broad uncertainty about methodological and statistical issues in multi-arm trials, leading to an increased likelihood of type I and II errors. The net effect on the results of the trials is unclear. Nonetheless, health research authorities, scientific journals and funders should therefore provide more explicit guidance to investigators about the appropriateness of adjustment or control of type I error rate in multi-arm trials and this should include an assessment of whether a multi-arm trial is the most efficient strategy for specific research question. The decision tool employed in this chapter may represent an important way forward in considering the need to type I error adjustment.

Of interest, industry funded multi-arm trials were less likely than non-industry funded multi-arm trials to report details about planned type I error adjustment in the study protocol when it was needed based on the decision tool. This finding is limited by the small number of non-industry funded trials in this thesis chapter and should be interpreted cautiously. However, it may be that there is more potential for selective reporting of statistical analyses in industry funded multi-arm trials compared to non-industry funded trials.

Although the focus of this thesis chapter was on trial protocols in the combined ASPIRE and Swiss cohort, an analysis of published RCTs from the ESORT cohort was also

performed. The results of the ESORT analysis were often similar to the ASPIRE and Swiss cohort but an important difference was in the prevalence of adjustment for multiple testing in trials where it was not needed based on the decision tool. In the ESORT analysis, 4 of 12 trials (33%, 95% CI 13%-62%) that did not require an adjustment of the type I error rate performed an adjustment and 1 trial (8%, 95% CI 1%-41%) used an alternative method to control the type I error rate. In contrast, only 1 of 11 trial protocols (9%, 95% CI 1%-44%) in the ASPIRE and Swiss cohort intended to adjust the type I error rate when it was not needed. The reason underlying the discrepancy between published RCTs and trial protocols cannot be completely elucidated from this study but it may be that adjustments for multiplicity may be added during peer review of manuscripts. For instance, in an analysis of 93 published RCTs in BioMed Central medical journals, peer reviewer comments resulted in unplanned additional analyses in 15 trials (16%) that were not pre-specified in the trial protocol (19). In the absence of clear recommendations regarding the adjustment of the type I error rate in multi-arm trials, researchers may be limited in their rebuttal to reviewers and may adjust their analysis plan. At the time of preparation of this thesis in August 2017, only 3 of the multi-arm trials in the ASPIRE cohort were published and data extraction was still ongoing for the Swiss cohort. This prohibited a robust analysis comparing trial protocols and publications. Further research is needed to better clarify how multi-arm trials are ultimately analyzed relative to their pre-specified statistical analysis plan.

Finally, the analysis of the ESORT data demonstrated that multi-arm trials were inadequately reported in the final publication. Indeed, 49 of 133 multi-arm trials (37%) did not report a sample size calculation. This finding is consistent with existing studies. For instance, investigators conducted a review of 221 published multi-arm trials and noted that

62 trials (28%) reported a sample size calculation (3). Information on sample size calculation is important to understand whether the trial was sufficiently powered for the pre-specified comparisons among treatment arms. Improving the reporting of multi-arm trials is therefore an important step to allow readers to ascertain the validity of the findings within multi-arm trial.

Implications for Clinicians, Policy Makers and Future Research

A potentially effective strategy to improve the design and reporting of multi-arm trials is through the development of guidelines. Currently, there are no guidelines for the methodological items that should be included in clinical trial protocols for multi-arm trials and an extension to the SPIRIT checklist may be warranted. Assuming more widespread adoption of the SPIRIT guidelines, the development of a multi-arm trials extension would directly address this gap. A draft list of items for consideration in a SPIRIT extension is provided in Table 6.6. The items were chosen based on the findings of this chapter.

Of interest, as of August 2017, a CONSORT extension for multi-arm trials is currently under development and it will be important to examine whether publication of these guidelines is associated with improvements in reporting over time.

Table 6.6: Potential Items for Inclusion in SPIRIT Extension for Multi-Arm Trials

SPIRIT Checklist Item	Phase I Trial Specific Extension
Introduction	
Item 6a: Background and Rationale	The Rationale for each intervention arm should be explicitly detailed
Item 7: Objective	Each intervention arm in a multi-arm trial should be explicitly mentioned in the primary study objectives
Item 7: Trial Design	The allocation ratio to each study arm should be explicitly reported. If unequal

	allocation is used, the reason underlying this decision should be detailed.
	The rationale for selecting a multi-arm trials as opposed to other trial designs should be detailed.
Methods: Participants, Interventions, Outcomes	
Item 14: Sample Size	If a type I error rate adjustment is performed, it should be explicitly stated whether the sample size will be adjusted to account for the new type I error rate
Methods: Data Collection, Management, Analysis	
Item 20a: Outcomes	Each intervention arm should be explicitly mentioned in the primary analysis plan for the primary outcome
	The analysis plan should detail whether any adjustment of the type I error rate will be performed and the reason for this decision.
	If a type I error rate adjustment is performed, it should be explicitly stated whether a global test will be performed to inform any subsequent comparison

Strengths and Limitations

The most important limitation in this analysis is the small number of trial protocols for multi-arm trials in the ASPIRE cohort. However, to increase the sample size, this chapter was conducted collaboratively with other investigators assessing the methodological conduct of clinical trials in Switzerland. Furthermore, an analysis of published RCTs in the ESORT cohort was also conducted. This larger cohort allowed for a more robust assessment of multi-arm trials.

Conclusion

Taken together, important gaps exist in the planned analytic strategy for multi-arm trials, that may have implications for the results obtained. Guidelines for the preparation of trial protocols for multi-arm trials may be beneficial.

7

RESULTS

THEME: SPECIAL TOPICS

Assessment of the Design, Conduct, and Dissemination of Phase I Trials

Drug development involves multiple stages of testing, in which drugs are sequentially assessed in different patient populations to establish their safety and their efficacy. After a drug has been demonstrated to be safe and efficacious in animals, there are four phases of clinical trial assessments required before widespread use in humans (1).

Phase I trials are defined as trials of novel therapeutic agents that are tested in a small population of healthy adults or adults with the health condition of interest. The intent of these trials are non-therapeutic and the aim is to demonstrate the safety, tolerability and pharmacokinetics of novel agents in a small population of adults (1,2). This can include estimating tolerable dose ranges for a drug, assessing drug-drug interactions or obtaining early estimates for drug activity. Phase I studies may be NRTs or RCTs and they are often conducted without blinding of study participants. In phase II trials, drug development focuses on assessments of efficacy, in a larger population of

healthy adults or adults with the medical condition of interest (1,2). An additional goal is to determine the appropriate dosing strategy for the drug and to assess the ascertainment of end-points for phase III trials. Doses in phase II trials are often lower than the highest dose used in phase I trials. Phase II trials may also be NRTs or small RCTs.

Phase III studies are conducted in adults with the condition of interest, largely to confirm the efficacy of the novel therapeutic agent and to demonstrate clinically significant effects (1,2). The utility of drugs in wider patient populations may also be considered, as well as the impact of the drug when combined with an additional agent (1,2). Drug approval is often based on the results of phase III studies, with consideration of results from the earlier phase studies. Phase III studies are often large multicentre double blind RCTs. Phase IV studies are largely intended to assess the efficacy and safety of novel agents after approval (1,2). They can also be conducted to support use of the drug for other treatment indications. The number of trials within each phase, as well as the methodological quality of the trials in each phase may vary based on the specific medical condition under study.

Given the importance of phase II-IV trials as studies that are used in regulatory approval of novel therapeutics, there is extensive evidence about the design, conduct, completion and dissemination of these trials (3-7). However, phase I trials have been largely overlooked but a similar amount of rigour is required when conducting these trials, particularly as the risk associated with these agents is often poorly delineated. The results of the trials, particularly related to adverse events, should also be shared to inform the development of future studies.

Although uncommon, oversights in the design and conduct of phase I trials have resulted in long-term morbidity and mortality. One of the first well publicized cases of serious morbidity and mortality in a phase I trial was in a gene therapy study in 1999 in the US. In this ascending

dose study of 18 participants, the last participant, Mr. Jesse Gelsinger, died after being administered the study drug (8-10). Multiple concerns were raised regarding the consent process and ethics review for the trial (10). An important methodological concern was also raised with respect to the trialists' decision to continue increasing the dose of the study drug despite signs of toxicity being demonstrated in earlier trial participants (10).

Two years later, another serious adverse event (SAE) – which is defined by the US FDA "any adverse event or adverse reaction that results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity, or is a congenital anomaly or birth defect" – occurred in a 2001 US study of hexamethonium, a drug which was previously used to treat high blood pressure (8,9). In this trial of healthy participants, Ms. Ellen Roche, developed a strong reaction to the medication, ultimately leading to cardiovascular shock and death (8,9). Two important gaps were identified during regulatory review of this study following the adverse event. First, regulatory authorities discovered that similar trials of hexamoethonium had been conducted and SAEs occurred among these participants (8,9). However, the results of the study were never published. Second, the researchers coordinating the hexomethonium study did not conduct a detailed systematic review of the literature when developing their study and failed to acknowledge previous studies demonstrating well-established risks associated with the use of hexomethonium (8,9). This led to a poorly informed risk assessment by the RECs approving the study.

Despite extensive efforts to improve the regulation of phase I trials following the deaths of Mr. Jesse Gelsinger and Ms. Ellen Roche, events leading to long term morbidity and mortality have recurred in phase I trials. For instance, in a UK phase I study of the monoclonal antibody TGN1412, administration of this agent to healthy participants resulted in life threatening toxicities

in six volunteers (11). In subsequent investigations, three central issues were identified in the conduct of the trial: first, the starting dose of TGN1412 was overestimated; second, there was only a 10-minute observation period to assess for adverse events between dosing of the first participant and subsequent participants; and third, a related drug showed similar complications in humans a decade earlier but the results of this study were never published (12,13). The events in the TGN1412 study led to the commissioning of a review of the conduct of phase I trials in the UK. This review produced the Duff report, which included 22 recommendations to improve the conduct and oversight of phase I trials (12). The key recommendations focused specifically on the issues of estimation of starting doses and intervals of observation. Other societies released comparable reports (14), reinforcing the importance of the Duff report.

Nonetheless, in 2015, a phase I ascending dose trial of the drug BIA 10-2474 led to the death of one healthy volunteer and the hospitalization of four others in France. The underlying methodological problem in this trial related to uncertainty about the interval of observation in the study (15,16). Although the study drug was administered sequentially in the first cohort of the ascending dose study with a 24-hour interval of observation period, the subsequent cohorts were not dosed with an observation period between participants (15). Therefore, multiple participants in the second highest ascending dose cohort had already been dosed by the time the first participant began experiencing an adverse drug reaction (15). While the SAE in the first participant may not have been prevented, the events in subsequent participants could readily have been prevented if a well justified interval of observation was specified for all trial cohorts.

Despite the well publicized nature of these SAEs, it is unclear whether these recurrent methodological issues reflect a broader problem in the design of phase I trials; in large part because few phase I trials are published (2). Indeed, an analysis of phase I trials approved by RECs in

France demonstrated that only 21 of 127 (17%) completed phase I trials were published, compared to 93 of 218 (43%) phase II-IV trials (2). Although there have been calls for a centralized system to monitor SAEs in phase I trials, regulatory authorities have been hesitant to act due to concerns about the commercial sensitivity of phase I trials. Empirical evidence to guide more robust regulatory oversight of phase I trials is therefore needed. Accordingly, this study was conducted to analyze the design, conduct and dissemination of phase I trials, with comparison of the dissemination to phase II-IV trials.

7.1 CHAPTER SPECIFIC METHODS

Definitions

Phase I Trials: Trials of novel therapeutic agents that are tested in a small population of healthy adults or adults with with the health condition of interest. The intent of these studies are non-therapeutic and the aim is to demonstrate the safety, tolerability and pharmacokinetics of novel agents in a small population of adults. For the purposes of this thesis, trial phase was determined based on the stated objectives of the study, and/or details in the study protocol or IRAS form specifying the phase of the trial. All IRAS applications were required to specify the phase of their trial. An example of a phase 1 trial is a trial to assess the safety and immunogenicity of a novel flu vaccine in healthy adults.

Serious Adverse Event: Any adverse event or adverse reaction that results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity, or is a congenital anomaly or birth defect.

Data Extraction

For phase I and phase II-IV trials, information on the general characteristics of the trials was extracted, as described in the Methods chapter. Biologicals were defined in accordance with the FDA guidelines as "virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, or analogous product, … applicable to the prevention, treatment, or cure of a disease or condition of human beings" (17). This includes monoclonal antibodies; cytokines such as interferons; enzymes such as thrombolytic agents; immunomodulatory drugs; and growth factors that alter hematopoiesis.

Information on whether the trials were registered in a trial registry was also extracted. If a trial registration number was not available in the ethics application, a search of trial registries was conducted, as described in the methods of Chapter Five. Briefly, a systematic search was conducted of the WHO International Clinical Trials Registry Platform, the European register of clinical trials, clinicaltrials.gov and the ISRCTN. As recommended by bibliographic studies (10), the search was conducted on the basic search interface for all registries using the following search strategy: the title of the protocol; the name of the principal investigator and a combination of keywords (e.g. the intervention name, health condition and trial identifier). The search for all trials was completed in April 2016. Notably, European Union (EU) legislation does not require that trial registry entries for phase I trials are publicly visible (18). Therefore, when the trial registration number provided in the ethics application is entered into the EU clinical trials register, there are

no records retrieved and the user is told that the trial cannot be found. When this occurred, the study was identified as "registered without a publicly accessible registry".

Finally, investigators' responses to questions in the IRAS form about results dissemination were assessed. Specifically, investigators must answer: "*How do you intend to report and disseminate the results of the study? Tick as appropriate:*

peer reviewer scientific journals;

- *internal report; conference presentation;*
- *publication on website; other publication;*
- *submission to regulatory authorities;*
- *access to raw data and right to publish freely by all investigators in study or by independent steering committee on behalf of all investigators;*
- *no plans to report or disseminate the results;*
- *other, please specify*"

For first-in-human phase I trials, guidelines produced after the TGN1412 phase I trial provide specific recommendations for the estimation of starting doses of biological agents and for the method of administering medications and monitoring for adverse events (12,19). With respect to estimation of the starting dose, the Minimum Anticipated Biological Effect Level (MABEL) or Pharmacologically Active Dose (PAD) is recommended over the No Observed Adverse Effect Level (NOAEL). This is because the MABEL estimates the minimum threshold for any biological activity whereas the NOAEL estimates the minimum threshold for harm. The MABEL and NOAEL are derived from testing of the novel agent in animal models and may be identical for some investigational medicines. However, they can also be remarkably different and the lowest

estimated starting dose should be used. The MABEL may not be readily calculated in every trial but it is recommended that it should be discussed. With respect to the method for administering medications, sequential dosing is recommended, which is when the first participant is dosed and is monitored for adverse events for an explicitly justified period of observation before the next participant is dosed. Details from first-in-human studies related to these methodological elements were extracted.

Upon completion of a trial, investigators are required to submit a declaration of end of study form and a end of study report (required within 1 year of completion) to the REC. For phase I trials that were completed, details on the date of study completion and whether it was terminated early were extracted from the end of study documents. If an end of study report was submitted to the committee, details on the number of adverse events leading to study withdrawal as well as the number of SAEs were extracted. SAEs were defined as "any adverse event or adverse reaction that results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity, or is a congenital anomaly or birth defect" (20). Information on whether the adverse event was related to the investigational medicine was also extracted. The rates of adverse events leading to study withdrawal and SAEs were the primary focus of this study because these events were consistently reported in all end of study reports.

Finally, PubMed and trial registries were searched in November 2016 for publications or dissemination of the results of the completed phase I studies that submitted their end of study form to the REC. The search strategy incorporated a combination of keywords, investigator names, intervention names and trial registration numbers related to the phase I trial. Additionally,

published phase II-IV studies from the ASPIRE cohort were identified and the rate of publication or results dissemination in these trials was compared to that of the phase I trials.

Data Analysis

A descriptive analysis was performed, as detailed in the Methods chapter. Multilevel Poisson regression with robust errors was used to assess the association between trial phase (phase I versus phase II-IV) and any dissemination of trials results, defined as either publication of the results in a scientific journal or dissemination of the results in a trial registry. The model was unadjusted.

7.2 CHAPTER RESULTS

General Characteristics of Phase I Trials

The study cohort was composed of 42 phase I trials and 217 phase II-IV trials (Figure 7.1). Phase I trials were more often single centre, industry funded and smaller in sample size compared to phase II-IV studies (Table 7.1). There was no difference in the proportion of oncology trials between phase I and phase II-IV trials (Table 7.1). Specific to phase I studies, 17 (40%) of 42 trials were first in human studies and 23 (55%) were limited to healthy participants. Half of all phase I studies were ascending dose studies and involved the use of biological agents (Table 7.1).

Figure 7.1: Identification of Studies in the Analysis of Phase I Versus Phase II-IV Trials

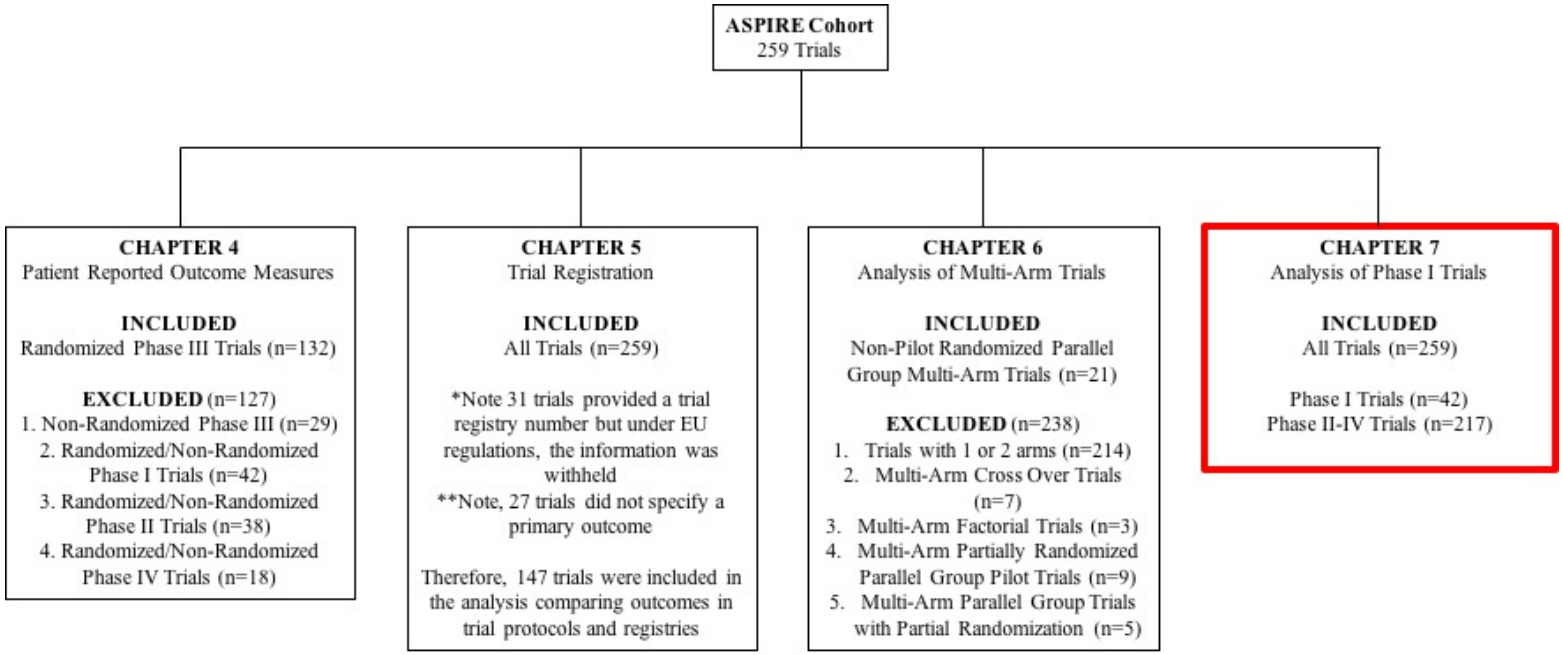


Table 7.1: General Characteristics of Phase I Trials and Phase II-IV Trials (n=259)

	Phase I (n=42)	**Phase II-IV (n=217)**	**P-value**
General Items			
Number of Centres			
Single	24 (57)	81 (37)	0.051
Multiple	18 (43)	134 (62)	
Unclear	0 (0)	2 (0.9)	
Funding			
Industry Funded	36 (82)	93 (43)	<0.001
Non-Industry Funded	6 (14)	79 (36)	
Not funded or only in house source	0 (0)	24 (11)	
Unclear	0 (0)	1 (1)	
Not Reported	0 (0)	20 (9)	
Medical Specialty			
Oncology	11 (26)	42 (19)	0.31
Non-oncology	31 (74)	175 (81)	
Sample size			
Median (Inter-quartile interval)	37 (24-52)	96 (42-252)	<0.001
Items Specific to Phase I Studies			
First-in-human study			
Yes	17 (41)	-	
No	25 (59)	-	
Ascending Dose Study			
Yes	21 (50)	-	
No	21 (50)	-	
Healthy Participants Only			
Yes	23 (55)	-	
Mixed	2 (5)	-	
No	17 (40)	-	
Type of Agent			
Biological	21 (50)	-	
Non-biological	21 (50)	-	

Methodological Items Specific to Phase I Trials

Of the 17 first-in-human phase I trials, 14 (82%) involved the use of a biological agent (Table 7.2). With respect to the estimation of the starting dose in the trials involving a biological agent, two trials calculated the MABEL and four trials that did not calculate the MABEL provided a reason for not doing so. The remaining 8 trials (58%) did not discuss the MABEL in their approach for calculating the starting dose of the investigational medicine administered.

With respect to the dosage regimen, 12 of the 17 first-in-human studies (71%) specified that the investigational medicine would be sequentially administered to participants in the first cohort studied (Table 7.2). However, none of the 12 studies provided a justification for this interval of observation in the protocol. Where multiple ascending dose cohorts were examined within a first-in-human study, 7 of 11 studies (64%) incorporated sequential dosing for all cohorts.

Table 7.2: Methodological Characteristics of First-in-Human Phase I Trials (n=17)

	Number of Trials (%) [95% CI]
Ascending Dose Study	
Yes	13 (77) [51-91]
No	4 (24) [9-49]
Sequentially Administered Doses	
Yes	12 (71) [46-87]
No, reason provided	1 (6) [0.8-32]
No, no reason provided	0 (0) [NA]
Unclear	4 (24) [9-49]
If sequentially dosed (n=12)	
Median duration between dosing of first participant(s) and subsequent participants (days)	2 (1-3)
Justification for duration between dosing of participants	
Yes	0 (0) [NA]
No	12 (100) [NA]
Sequential dosing in all cohorts for ascending dose studies (n=11)	
Yes	7 (64) [34-86]
No	4 (36) [14-66]
If a biological agent was used (n=14)	

Was MABEL* used to estimate starting dose	
Yes	2 (14) [4-43]
No, reason provided	4 (29) [11-56]
No, no reason provided	8 (58) [32-79]

*MABEL is minimum anticipated biological effect level

Adverse Events in Completed Phase I Trials

Thirty-one of 42 studies were reported as completed to the REC based on the submission of an end of study form. Three trials (10%) were terminated early and of the remaining 28 studies, 27 were completed and should have submitted an end of study report to the ethics committee by the time of final data extraction for this chapter. However, only 21 studies (75%) provided this documentation and it was overdue for six studies.

In the 21 studies that were completed and submitted an end of study report to the REC, the total number of participants was 774 and the median achieved sample size was 27 (IQR 17-39). There were no deaths but a total of 10 SAEs occurred across five studies. Of these, eight events were unrelated to the investigational medicine administered.

The remaining two events that were potentially related to the investigational medicine in the study occurred in separate first-in-human trials of healthy participants. The first was an episode of asymptomatic atrial fibrillation which was noted on electrocardiogram after four doses of the study drug. The event was judged to be related to the study drug because of the temporality and the absence of other causes. The atrial fibrillation ceased after termination of the study drug. The second was an episode of rhabdomyolysis. This event was initially judged by the trial investigators to be related to the investigational medicine but this was later overturned by the principal

investigator. Adverse events leading to study withdrawal occurred in two studies. There were 10 events and all were due to nausea and vomiting.

Dissemination of Trial Information: All Trials and Completed Trials

All phase I trials (n=42) were registered in a trial registry but only 16 (38%) were publicly accessible as per EU regulations for phase I trials. Phase I trials were more likely than phase II-IV trials to be planned for submission to regulatory authorities but were less likely to be planned for publication in a scientific journal (Table 7.3). For the 17 phase I studies that were not planned for publication in a scientific journal, only two were intended to be published on a website and only one was intended to be presented at a conference (Table 7.3). Similar findings were noted for first-in-human studies (Table 7.3).

Twenty-eight of 42 phase I trials and 62 of 217 phase II-IV trials were completed as of November 2016 and were not terminated early. Five (18%) of 28 phase I trials reported results in the trial registry compared to 27 (44%) of 62 phase II-IV trials. However, 8 (29%) of 28 phase I trials were published compared to 19 (31%) of 62 phase II-IV trials. Overall, there was no significant difference in the frequency of any dissemination (publication as a scientific article or publication of results in a trial registry) between phase I and phase II-IV trials (risk ratio 1.32, 0.89-1.97). None of the trials with SAEs were disseminated.

Table 7.3: Trial Registration and Intended Dissemination Strategies for Phase I Compared to Phase II-IV Trials

ALL TRIALS	**Phase I (n=42) n (%) [95% CI]**	**Phase II-IV (n=217) n (%) [95% CI]**	**P-value**
Trial Registration			
Yes and publicly accessible	16 (38) [25-53]	103 (48) [41-54]	<0.001
Yes but not publicly accessible	26 (62) [47-75]	5 (2) [0.9-5]	
No	0 (0) [NA]	109 (50) [44-57]	
Intended for Submission to Regulatory Authority			
Yes	36 (86) [72-93]	85 (39) [33-46]	<0.001
No	6 (14) [7-28]	132 (61) [54-67]	
Planned Publication in Scientific Journal			
Yes	25 (60) [44-73]	210 (97) [93-98]	<0.001
No	17 (40) [27-56]	7 (3) [2-7]	
If not planned to be published in scientific journal, alternate dissemination method*	**(n=17)**	**(n=7)**	
Internal Report	7 (41) [21-65]	4 (57) [23-86]	-
Conference	1 (6) [0.8-32]	3 (43) [14-77]	-
Submission to Regulatory Authorities	15 (88) [63-97]	6 (100) [42-98]	-
Detailed on Website	2 (12) [3-37]	1 (14) [2-58]	-
FIRST IN HUMAN STUDIES (n=17)			
Trial Registration			
Yes and publicly accessible	3 (18) [6-43]	-	-
Yes but not publicly accessible	14 (82) [57-94]	-	-
No	0 (0) [NA]	-	-
Planned Publication in Scientific Journal			
Yes	11 (65) [40-83]	-	-
No	6 (35) [17-60]	-	-
If not planned to be published in scientific journal (n=6), alternate dissemination method*			
Internal Report	3 (50) [17-83]	-	-
Conference	0 (0) [NA]	-	-
Submission to Regulatory Authorities	6 (100) [NA]	-	-
Detailed on Website	1 (17) [2-63]	-	-
COMPLETED TRIALS	**(n=28)**	**(n=62)**	
Trial Registration			
Yes and publicly accessible	9 (32) [18-51]	39 (63) [50-74]	<0.001
Yes but not publicly accessible	19 (68) [49-82]	0 (0) [NA]	

No	0 (0) [NA]	23 (37) [26-49]	
Dissemination			
Any Dissemination	13 (46) [29-65]	38 (61) [49-73]	0.17**
Published in Scientific Journal	8 (29) [15-48]	22 (36) [25-48]	0.842
Results Posted to Trial Registry	5 (18) [8-36]	27 (44) [32-56]	0.002

*Note that studies may be counted more than once for this item **p-value taken from multilelvel poisson regression model

7.3 CHAPTER DISCUSSION

This study provides a detailed assessment of the design, conduct and dissemination of phase I trials. There were four major findings: First, less than half of the first-in-human studies involving biological agents discussed the MABEL in the estimation of the starting dose of their agent; Second, explicit justification for the interval of observation between sequential administrations of investigational medicines was not provided in any study; Third, SAEs were uncommon but did occur in phase I trials of healthy participants; Fourth, only 60% of phase I trials were intended to be published in a scientific journal, compared to 97% of phase II-IV trials; Finally, there was no difference in the dissemination of phase I trials compared to phase II-IV trials. However, none of the phase I trials with SAEs were disseminated.

Comparison with Prior Individual Studies

The events in the TGN 1412 and BIA 10-2474 trials have led to widespread calls for improvements in the oversight of phase I trials. However, the few published reviews of phase I trials provide only sparse details on the design and conduct of these studies. In the most comprehensive study to date, investigators examined protocols of phase I trials approved by

French RECs in 1994 but the authors only had access to information on the funding status, sample sizes, use of interim analyses and subsequent publication of the studies (2). Furthermore, the phase I trials examined in their study were dated (2), therefore preceding the TGN-1412 and BIA 10-2474 events as well as the establishment of trial registries. This analysis of the ASPIRE cohort therefore provides the first contemporary and comprehensive assessment of phase I trials and the findings highlight important aspects of the design, conduct and dissemination of phase I trials that may be amenable to improvement.

First, the MABEL is the recommended strategy for estimating starting doses in first-in-human phase I trials because it estimates the lowest threshold at which biological activity occurs. In the TGN 1412 study, the MABEL was not used or discussed and this led to a 10 fold overestimation of the starting dose in that trial, thereby contributing to adverse events (12). In this study, only 43% of first-in-human studies of biological agents used the MABEL to estimate their starting dose or discussed why it was not used.

An equally important methodological consideration is the interval of observation between sequential dosing of participants. Current recommendations are that investigators justify the time allowed, if any, between administration of an investigational medicine to the first and subsequent participants in a first-in-human phase I trial (19). In the event of an unexpected adverse event in the first trial participant, this period of observation ensures that subsequent participants are not unnecessarily exposed to the investigational medicine. None of the phase I trials in the ASPIRE cohort provided explicit justification for the interval of observation between sequential dosing of participants. While the absence of information on estimation of the starting dose or the interval of observation does not mean it was overlooked in the planning of the study, it is important for RECs to ensure that methodological considerations which have implications for patient safety are

explicitly detailed. Indeed, the interval of observation was not explicitly specified in the BIA 10-2474 trial protocol and this oversight contributed to the number of participants that experienced adverse events.

Consistent with previous studies (21-23), SAEs were uncommon in phase I trials examined in this study. There were two SAEs related to the investigational medicine among 774 participants. Although the study sample was small, this is the first examination the frequency of SAEs without restriction to a specific funder or contract research organization. Our independence from the actual conduct of the trials is also important because it reduces the potential for bias when reviewing study records. Notably, in the case of a SAE of rhabdomyolysis in a phase I trial in this study, it is clear that the judgement of causality can be subjective and there may be uncertainty in the decisions made by investigators. If necessary, the centralized monitoring can be performed in confidential manner for a select group of phase I trials.

Implications for Clinicians, Policy Makers and Future Research

The dissemination of information regarding phase I trials was similar to phase II-IV trials; a finding which is contrary to existing studies (2). These results are notable because a key argument against the establishment of a centralized registry to monitor SAEs in phase I trials is that the results of the trials are commercially sensitive and the intellectual property of the trialists and sponsors must be protected. A centralized and confidential system to record SAEs could be established, which would therefore allow RECs to improve their apriori risk assessment of phase I trials, while also protecting the commercial interest of research investigators. To date, there has been no progress on centralized data collection for phase I trials, and in the UK, the NRRES faced

litigation when it introduced requirements for the sharing of phase I trial results (24). Our findings reinforce the need for ongoing efforts to improve the sharing of the results of phase I trials.

An additional strategy to improve the design and conduct of phase I trials is through the development of guidelines for phase I clinical trial protocols. Research regulatory authorities, such as the NRES in the UK, are considering adoption of the SPIRIT guidelines. However, as of July 2017 the current iteration of the SPIRIT guidelines does not address important methodological items related to phase I trials, including strategies for the estimation of starting doses and justification for intervals of observations. Assuming more widespread adoption of the SPIRIT guidelines, the development of a phase I trials extension would directly address these gaps. A draft list of items for consideration in a SPIRIT extension is provided in Table 7.4. The items were chosen based on the findings of this chapter as well as draft guidance prepared by the EMA for reducing risks in first-in-human trials following the BIA-10-2474 trial.

Table 7.4: Potential Items for Inclusion in SPIRIT Extension for Phase I Trials

SPIRIT Checklist Item	**Phase I Trial Specific Extension**
Administrative Information	
Item 2: Trial Registration	Details should be provided as to if and/or when the trial registry entry will be publicly accessible
Introduction	
Item 8: Trial Design	Explicit details should be provided as to whether the trial is a first in human study
	Explicit details should be provided as to whether the trial is an ascending dose study
Methods: Participants, interventions, and outcomes	
Item 11: Interventions	Method for estimation of starting dose
	If a biological agent is used, the MABEL should be used to estimate the starting dose or a reason should be provided.
Item 13: Participant Timeline	If the trial is a first in human study, explicit details about whether participants will be

	dosed sequentially should be provided. If dosing is not performed sequentially, justification should be provided.
	If the trial is a first in human study with an ascending dose design, explicit details about whether all ascending dose cohorts will be dosed sequentially should be provided. If dosing is not performed sequentially, justification should be provided.
	For first-in-human trials with sequential dosing, the duration of the interval of observation between dosing of the first participant and subsequent participant should be explicitly stated AND explicitly justified. If no period of observation, justification should be provided.
	For first-in-human trials with, the duration of the interval of observation between dosing of the first participant and subsequent participant should be explicitly stated AND explicitly justified. If no period of observation, justification should be provided.
Ethics and Dissemination	Plans for dissemination of details about serious adverse events if the study will not be published in a scientific journal

Strengths and Limitations

The strength of this study is that it includes an unselected cohort of clinical trials from a variety of sponsors and medical specialties. This enhances the generalizability of the findings to other jurisdictions. Nonetheless, some limitations should be considered when interpreting the results of this study. First, end of study reports were missing for six studies because they were overdue for submission to the REC. As human resources at the REC are limited, regular follow up with researchers who do not adhere to these policies can be delayed. Nonetheless, the inclusion of these studies is unlikely to change the interpretation of the results in this thesis chapter. Second, the sample size in this study was small but the analysis provides the most detailed assessment to date of the conduct of phase I trials. Further work, specifically focused on first-in-human studies is needed to better examine the methodological conduct of these studies.

Conclusion

Taken together, phase I trials are generally safe but there are important opportunities to improve the design, conduct and dissemination of these studies. Addressing these methodological gaps and improving monitoring of SAEs may have implications for patient safety and for the planning of future phase I trials.

8

CONCLUSION

Taken together, this thesis is a detailed examination of different methodological issues related to clinical trial conduct. This was possible through unprecedented access to study protocols for clinical trials approved by 19 RECs in the Bristol office of the UK NRES. As well, this thesis benefited from access to detailed correspondence between the RECs and clinical trial investigators, which allowed for the assessment of trial results and the ascertainment of adverse events. The findings of this thesis provide insights into how the design, conduct and dissemination of clinical trials can be improved.

To briefly review, Chapter One of this thesis introduced clinical trials as a key source of healthcare evidence. The chapter also highlighted the importance of considering the methodological design and characteristics of clinical trials when interpreting the results. Chapter Two detailed the methods that were employed in this thesis, particularly with respect to the use of multilevel models and poisson regression to generate estimates of the association between general trial characteristics and different methodological outcomes.

Chapter Three introduced the ASPIRE cohort. The objectives of this chapter were three-fold. First, it was to provide an overview of the clinical trials included in the ASPIRE cohort. Second, it was to assess the quality of reporting in clinical trial protocols using the items specified in the SPIRIT checklist. Third, it was to determine what study characteristics were associated with more frequent reporting of important methodological items. Overall, the ASPIRE cohort was made up of 259 clinical trials and was unique among other studies of trial protocols because of the inclusion of NRTs as well as early phase trials. Many of the clinical trials in the ASPIRE cohort were multicentre trials involving sites in the European Economic Area and internationally. Findings based on the ASPIRE cohort were therefore considered to be generalizable to other jurisdictions. Using the SPIRIT checklist, Chapter Three also demonstrated that clinical trial protocols were variable in the reporting of a range of items that are central to good trial conduct. Indeed, important items such as random sequence generation and allocation concealment were reported in fewer than 80% of clinical trial protocols. In multivariable regression analyses, there were no observable associations between general trial characteristics and the reporting of a range of methodological items in clinical trial protocols. However, the sample sizes were small and confidence intervals were wide. These findings suggest that there is large potential for improvement in the preparation of clinical trial protocols. The adoption and use of the SPIRIT checklist may be one strategy through which this could be achieved.

Chapter Four examined the use of PROMs in phase III RCTs in the ASPIRE cohort. The objectives of this chapter were to determine what proportion of clinical trial protocols included PROMs and to assess the inclusion of PROMs based on trial characteristics such as the type of outcome, intervention and disease area. The chapter also sought to determine whether methodological challenges related to the use of PROMs – validity, sample size calculations and

missing data – were accounted for in trial protocols. One-hundred and thirty-two phase III RCT protocols were included in the analysis for Chapter Four. Among these, 94 (71%) included PROMs. There were also no study characteristics that were clearly associated with underuse of PROMs. Finally, key methodological items that are pertinent to the use of PROMs as primary outcomes – validity of PROMs, strategies for missing data and components of sample size calculations – were infrequently addressed in clinical trial protocols; oversights that have the potential to bias estimates obtained from clinical trials involving PROMs. Efforts are therefore needed to improve how PROMs are incorporated into clinical trials. The development of an extension to the SPIRIT guideline, which focuses on methodological items specific to PROMs may be one way to guide authors considering the use of PROMs in their clinical trial. The appropriate use of PROMs in clinical trials can also be ensured during the research ethics review process, particularly if RECs focus on methodological issues related to PROMs.

Chapter Five focused on trial registration, given its importance as a strategy to improve accountability in the conduct and reporting of clinical trials (1). The objectives of the chapter were three-fold. First, it was to assess the frequency of trial registration. Second, it was to assess the fidelity of information in trial registries by comparing the primary and secondary outcomes in clinical trial protocols to the trial registry entry. Third, it was to determine reasons for which trials were not registered based on detailed documentation in the IRAS form. There were 259 clinical trials analyzed in this chapter, of which 205 (79%) were registered in a trial registry and 54 (21%) were not registered. Concordance between the trial protocol and trial registry was modest for the classification of the study outcome and the definition of the measurement variable. However, concordance between the two sources was poor for the time point of the study outcome and for the analytic metric. Of interest, pharmacologic trials and multicentre trials were less likely to have

discrepancies between the protocol and registry as compared to non-pharmacologic and single centre trials, respectively. Finally, common reasons for non registration of trials included reluctance to register pilot studies or trials of behavioural interventions, as well as misconceptions about costs related to trial registration. The analyses conducted in this chapter provide insight into whether trial registration will ultimately impact study findings. Indeed, if the trial registry is not an accurate reflection of the trial protocol, its utility in ensuring accountability for the conduct of the trial is undermined. Furthermore, given the extent of discrepancies between trial protocols and trial registries, it is difficult to anticipate that any regulatory research authority could regularly ensure the fidelity of information entered into the trial registries. It may therefore be pragmatic to ensure researchers publish their study protocols. This would allow other researchers, clinicians and the general public to have a clearer sense of how study outcomes have changed during the corse of the trial. Finally, in order to ensure all trials are registered, a concerted effort is needed to dispel myths related to trial registries. This effort should be targeted at researchers and journal editors, both of whom hold misconceptions about trial registration.

Chapter Six was focused on the epidemiology and design of multi-arm trials. Multi-arm trials make up a modest share of the clinical trials that are conducted (2,3). While there are important benefits associated with multi-arm trials, including increased efficiency and potentially faster recruitment, these trials may be statistically more complex (4). As information on multi-arm trials is limited to assessments of published articles, this chapter sought to provide more insight into how these trials are designed. The objective of the chapter was to characterize the planned design and analysis of multi-arm trials. Notably, multi-arm trials made up 11% of RCTs in the ASPIRE cohort, but this percentage was consistent with the Swiss cohort of clinical trial protocols as well as a cohort of RCTs published and indexed in PubMed in December 2012. Most multi-arm

trial protocols in the combined ASPIRE and Swiss cohort accounted for all study arms in their objectives, sample size calculation and planned analysis. However, when a type I error adjustment was needed based on a published decision tool, few trials stated whether the type I error rate would be controlled. Furthermore, only half of the trials that planned to adjust their type I error rate also adjusted their sample size calculation to accommodate for the reduction in statistical power. These latter items may have implications for the results of the trial. Given the complexity of multi-arm trials, the development of an extension to the SPIRIT guidelines that specifically target these methodological issues would be important.

Finally, Chapter Seven examined the design, conduct and dissemination of phase I trials. Phase I trials involve the early testing of novel therapeutic agents in humans. As the risk associated with these agents is often poorly delineated, it is important that phase I trials are carefully conducted. The results of these trials, particularly related to adverse events, should be shared to guide future studies. However, there is limited empirical evidence on the characteristics and results of phase I trials, in part because so few are published. There were 42 phase I trials and 217 phase II-IV trials in the ASPIRE cohort. Among first-in-human phase I trials, less than half of the studies involving biological agents discussed recommended strategies in the estimation of the starting dose of their agent. Furthermore, explicit justification for the interval of observation between sequential administrations of investigational medicines was not provided in any first-in-human phase I study. Serious adverse events were uncommon but did occur in phase I trials of healthy participants. Finally, although, phase I trials were equally as likely to be published as phase II-IV trials, none of the phase I trials with SAEs were published. The results of Chapter Seven have clear implications. First, the development of a SPIRIT extension specific to methodological issues in phase I trials may be warranted, particularly as these methodological issues have implications for

patient safety. Furthermore, the non-publication of phase I trials is particularly problematic when adverse events occur. A centralized system for monitoring adverse events may allow for more robust apriori risk assessment of phase I trials by RECs and also inform the design of future phase I trials.

STRATEGIES TO OVERCOME LIMITATIONS IDENTIFIED IN THIS THESIS

The most consistent limitation in this thesis is the sample size. This was particularly important for the results obtained in Chapter Four (PROMs), Chapter Six (multi-arm trials) and Chapter Seven (phase I trials). In Chapter Four, the sample size was small and in regression analyses, the confidence intervals were wide. In Chapter Six and Chapter Seven, the number of trials analyzed was small, therefore limiting the robustness of the findings. A formal sample size calculation was not performed for the ASPIRE study. However, the focus was on ensuring a sufficient number of protocols were obtained for estimation of the baseline quality of reporting of clinical trial protocols. Given that the majority of this thesis focused on research questions that involved specific groups of trial protocols, as opposed to the entire cohort, the focus of a sample size discussion should have been on achieving a sufficient number of trial protocols for the smallest group examined (e.g. 50 trial protocols for the study on multi arm trials). A crude sample size calculation can be performed based on the anticipated number of multi-arm trials. In reviews of published RCTs, the frequency of multi-arm trials was 23%. If a conservative estimate for the proportion of multi-arm trials among all trial (published and unpublished RCTs and NRTs) was empirically assumed to be 10%, the number of trial protocols that would be required to obtain at least 50 multi-arm trial protocols would be 500 trial protocols. This sample size would have

allowed for a more robust assessment of the characteristics and analysis of multi-arm trials as well as more robust analyses related to PROMs and early phase trials.

FUTURE RESEARCH QUESTIONS

Nonetheless, this thesis serves as the foundation for many other studies based on the ASPIRE cohort. First and foremost, an important aim of this thesis was to provide a baseline assessment of the quality of reporting in clinical trial protocols prior to the introduction of the SPIRIT checklist in 2013. This assessment would serve as a comparator for subsequent assessments of clinical trial protocols approved after the introduction of the SPIRIT checklist and would allow for an evaluation of the impact of SPIRIT on the quality of clinical trial protocols. It should be noted that an interrupted time-series analysis may be the best analytical approach to evaluate the impact of SPIRIT as opposed to a before-and-after analysis. In a before-and-after analysis, the comparison would be made between two instantaneous points in the time prior to and after the introduction of the SPIRIT guidelines. In contrast, a time series analysis allows for evaluating the effect of the SPIRIT guidelines while accounting for any trends in the quality of reporting in clinical trial protocols over a time period prior to its introduction in 2013. Time-series analyses are improved if multiple time points are included within the regression model. Therefore, cohorts of clinical trial protocols should be sampled for at least 3 time points (e.g. 2014, 2015, and 2016) following the introduction of the SPIRIT guideline. Conducting a time-series analysis was beyond the scope of this thesis because access was only provided to clinical trial protocols approved in 2012.

The next logical step in work related to the ASPIRE cohort is to examine the publications for each of the trial protocols. Specially, there is an opportunity to build on research questions already examined in this thesis. For instance, Chapter Three assessed the reporting of SPIRIT checklist items in trial protocols. An important research question would be to assess which SPIRIT checklist items are omitted or newly introduced in the trial publication. This analysis could conceivably focus on methodological items and general characteristics associated with bias: random sequence generation, allocation concealment, blinding and funding source. Chapter Four examined the use of PROMs and noted that there were important gaps with respect to methodological items that are relevant to PROMs. An important research question extending upon this work is to compare how these methodological items are addressed in the trial publication and whether there are any differences compared to the trial protocol. Chapter Five compared the reporting of outcomes in trial protocols and trial registries. Building on this work, it would be particularly interesting to determine whether discrepancies between the trial protocol and trial registry persist in the published article. Furthermore, it would be novel to assess whether discrepancies between the two sources are associated with statistically significant results for the study outcome. Although Chapter Six and Seven have already included an assessment of the published article, the sample size in these chapters was small. Therefore, it would be important to repeat these studies with a larger sample size. It might also allow for the comparison of methodological items in these chapters based on study characteristics such as funding source.

FINAL REMARKS

Taken together, this thesis has provided insight into different methodological issues in the design, conduct and dissemination of clinical trials. It has also created a foundation for future work in this area based on the ASPIRE cohort. It is hoped that the insight garnered from this thesis as well as any future work will inform the development of guidelines regarding clinical trial design and reporting and will inform research ethics review of clinical trials.

LIST OF REFERENCES

REFERENCES CHAPTER ONE

1. Hennekens CHBJE, Mayrent SL. Epidemiology in medicine. Boston : Little Brown and Company, 1987; 1987.

2. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004 May 26;291(20):2457–65.

3. Pildal J, Chan A-W, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ. 2005 May 7;330(7499):1049.

4. Lundh A, Krogsboll LT, Gøtzsche PC. Sponsors' participation in conduct and reporting of industry trials: a descriptive study. Trials. Trials; 2012 Aug 24;13(1):1–1.

5. Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleza-Jeric K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. 2013. pp. 200–7.

6. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ. 2013;346:e7586.

7. Danielsen AK, Okholm C, Pommergaard H-C, Burcharth J, Rosenberg J. Number of published randomized controlled multi center trials testing pharmacological interventions or devices is increasing in both medical and surgical specialties. PLoS ONE. 2014;9(7):e101383.

8. Ahmed Ali U, van der Sluis PC, Issa Y, Habaga IA, Gooszen HG, Flum DR, et al. Trends in Worldwide Volume and Methodological Quality of Surgical Randomized Controlled Trials. Annals of Surgery. 2013 Aug;258(2):199–207.

9. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919.

10. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010 Mar 23;340(mar23 1):c723–3.

11. Downing NS, Shah ND, Neiman JH, Aminawung JA, Krumholz HM, Ross JS. Participation of the elderly, women, and minorities in pivotal trials supporting 2011-2013 U.S. Food and Drug Administration approvals. Trials. 2016 Apr 14;17:199.

12. Cherubini A, Oristrell J, Pla X, Ruggiero C, Ferretti R, Diestre G, et al. The persistent exclusion of older patients from ongoing clinical trials regarding heart failure. Arch Intern Med. 2011 Mar 28;171(6):550–6.

13. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. JAMA. 2004 Jun 9;291(22):2720–6.

14. Coca SG, Krumholz HM, Garg AX, Parikh CR. Underrepresentation of renal disease in randomized controlled trials of cardiovascular disease. JAMA. 2006 Sep 20;296(11):1377–84.

15. Konstantinidis I, Nadkarni GN, Yacoub R, Saha A, Simoes P, Parikh CR, et al. Representation of Patients With Kidney Disease in Trials of Cardiovascular Interventions: An Updated Systematic Review. JAMA Intern Med. 2016 Jan;176(1):121–4.

16. Cochrane. Cochrane Handbook [Internet]. handbook.cochrane.org. Available from: http://handbook.cochrane.org/

17. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. New England Journal of Medicine. 2000 Jun 22;342(25):1887–92.

18. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. New England Journal of Medicine. 2000 Jun 22;342(25):1878–86.

19. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001 Aug 15;286(7):821–30.

20. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014 Apr 29;(4):MR000034.

21. Papanikolaou PN, Christidi GD, Ioannidis JPA. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. CMAJ. 2006 Feb 28;174(5):635–41.

22. Sterne JAC, Jűni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in "meta-epidemiological" research. Statistics in Medicine. 2002 Jun 15;21(11):1513–24.

23. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA. 1994 Jul 13;272(2):125–8.

24. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008 Mar 15;336(7644):601–5.

25. Savović J, Jones HE, Altman DG, Harris RJ, Jűni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012 Sep 18;157(6):429–38.

26. Nüesch E, Trelle S, Reichenbach S, Rutjes AWS, Bürgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ. 2009;339(sep07 1):b3244–4.

27. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. JAMA. 2014 Jan;311(4):368–77.

28. Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med. 2011 Jul 5;155(1):39–51.

29. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. Crit Care Med. 2009 Dec;37(12):3114–9.

30. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ. 2013;346:f2304.

31. Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2012 Dec 12;12:MR000033.

32. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ. 2003 May 31;326(7400):1167–70.

33. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. Lancet. 2004 Apr 24;363(9418):1341–5.

34. Hróbjartsson A, Boutron I, Turner L, Altman DG, Moher D. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge. Cochrane Database Syst Rev. 2013 Apr 30;(4):ED000058.

35. Chan A-W, Upshur R, Singh JA, Ghersi D, Chapuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. BMJ. 2006 May 6;332(7549):1086–9.

36. Chan A-W, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ. 2004 Sep 28;171(7):735–40.

37. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 1996 Aug 28;276(8):637–9.

38. Moher D, Jones A, Lepage L, CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. JAMA. 2001 Apr 18;285(15):1992–5.

39. To MJ, Jones J, Emara M, Jadad AR. Are Reports of Randomized Controlled Trials Improving over Time? A Systematic Review of 284 Articles Published in High-Impact

General and Specialized Medical Journals. Larivière V, editor. PLoS ONE. 2013 Dec 31;8(12):e84779.

40. Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. BMJ. 2015;350:h809.

REFERENCES CHAPTER TWO

1. HRA. Health Research Authority - Protecting and promoting the interests of patients and the public in health research [Internet]. hra.nhs.uk. [cited 2016 Nov 16]. Available from: http://hra.nhs.uk

2. HRA. Integrated Research Application System (IRAS) [Internet]. hra.nhs.uk. [cited 2016 Oct 28]. Available from: http://www.hra.nhs.uk/resources/applying-for-reviews/integrated-research-application-system-iras/

3. Chan A-W, Upshur R, Singh JA, Ghersi D, Chapuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. BMJ. 2006 May 6;332(7549):1086–9.

4. Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. Clin Trials. 2005;2(2):152–62.

5. Hennekens CHBJE, Mayrent SL. Epidemiology in medicine. Boston : Little Brown and Company, 1987; 1987.

6. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA. 1998 Nov 18;280(19):1690–1.

7. Knol MJ, le Cessie S, Algra A, Vandenbroucke JP, Groenwold RHH. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ. 2012 May 15;184(8):895–9.

8. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004 Apr 1;159(7):702–6.

9. Coutinho LMS, Scazufca M, Menezes PR. Methods for estimating prevalence ratios in cross-sectional studies. Revista De Saude Publica. 2008 Dec;42(6):992–8.

REFERENCES CHAPTER THREE

1. Hennekens CHBJE, Mayrent SL. Epidemiology in medicine. Boston : Little Brown and Company, 1987; 1987.

2. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, et al. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. Health Technol Assess. 2007 Nov;11(48):iii–ix–105.

3. Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan A-W. Constraints on publication rights in industry-initiated clinical trials. JAMA. 2006 Apr 12;295(14):1645–6.

4. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. 2014. p. g1687.

5. Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan A-W. Ghost authorship in industry-initiated randomised trials. PLoS Medicine. 2007 Jan 18;4(1):e19.

6. Flanagin A, Carey LA, Fontanarosa PB, Phillips SG, Pace BP, Lundberg GD, et al. Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. JAMA. 1998 Jul 15;280(3):222–4.

7. Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. BMJ. 2013 Mar 21;346(mar21 1):f1135–5.

8. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004 May 26;291(20):2457–65.

9. Chan A-W, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ. 2008;337(dec04 1):a2299–9.

10. Kasenda B, Schandelmaier S, Sun X, Elm von E, You J, Blumle A, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ. 2014 Jul 16;349(jul16 1):g4539–9.

11. Blumle A, Meerpohl JJ, Rucker G, Antes G, Schumacher M, Elm von E. Reporting of eligibility criteria of randomised trials: cohort study comparing trial protocols with subsequent articles. BMJ. 2011 Apr 5;342(apr05 2):d1828–8.

12. Kyte D, Duffy H, Fletcher B, Gheorghe A, Mercieca-Bebber R, King M, et al. Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. PLoS ONE. 2014;9(10):e110229.

13. Pildal J, Chan A-W, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ. 2005 May 7;330(7499):1049.

14. Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. BMC Medicine. 2015 Sep 15;13:221.

REFERENCES CHAPTER FOUR

1. Gnanasakthy A, Mordin M, Clark M, DeMuro C, Fehnel S, Copley-Merriman C. A review of patient-reported outcome labels in the United States: 2006 to 2010. Value Health. 2012 May;15(3):437–42.

2. Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. BMJ. 2010;341(nov01 1):c5707–7.

3. FDA. Guidance for Industry

Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims [Internet]. fda.gov. 2009 [cited 2015 Oct 2]. Available from: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf

4. Frank L, Basch E, Selby JV. The PCORI Perspective on Patient-Centered Outcomes Research. JAMA. 2014 Oct 15;312(15):1513.

5. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Contemporary Clinical Trials. Contemporary Clinical Trials. Elsevier Inc; 2015 Jul 1;43(C):1–9.

6. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006 Oct 11;4:79.

7. Carron M, Emery M-P, Marquis P, Piault E, Scott J. Recent Trends in the Inclusion of Patient-Reported Outcome (PRO) Data in Approved Drugs Labeling by the FDA and EMEA [Internet]. Vol. 40, pro-newsletter.com. 2008 [cited 2016 Nov 19]. pp. 8–10. Available from: http://www.pro-newsletter.com/images/PDF_articles/prolabels.pdf

8. Gnanasakthy A, DeMuro C, Clark M, Haydysch E, Ma E, Bonthapally V. Patient-Reported Outcomes Labeling for Products Approved by the Office of Hematology and Oncology Products of the US Food and Drug Administration (2010-2014). J Clin Oncol. 2016 Jun 1;34(16):1928–34.

9. DeMuro C, Clark M, Doward L, Evans E, Mordin M, Gnanasakthy A. Assessment of PRO label claims granted by the FDA as compared to the EMA (2006-2010). Value Health. 2013 Dec;16(8):1150–5.

10. EMA. Guideline on the evaluation of anticancer medicinal products in man [Internet]. ema.europa.eu. 2016 [cited 2016 Nov 20]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf

11. Izem R, Kammerman LA, Komo S. Statistical challenges in drug approval trials that use patient-reported outcomes. Stat Methods Med Res. 2014 Oct;23(5):398–408.

12. Evidence Matters [Internet]. [cited 2016 Nov 20]. Available from: http://www.nxtbook.com

13. Schandelmaier S, Conen K, Elm von E, You JJ, Blumle A, Tomonaga Y, et al. Planning and reporting of quality-of-life outcomes in cancer trials. Annals of Oncology. 2015 Sep;26(9):1966–73.

14. Kasenda B, Elm von E, You J, Blümle A, Tomonaga Y, Saccilotto R, et al. Prevalence, Characteristics, and Publication of Discontinued Randomized Trials. JAMA. 2014 Mar 12;311(10):1045.

15. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. BMJ. 2013 Oct 29;347(oct28 9):f6104–4.

16. Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg DAB. Issues in the registration of clinical trials. JAMA. 2007 May 16;297(19):2112–20.

17. Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. JAMA. 2008 Jun 4;299(21):2543–9.

18. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. JAMA. 2014 Jan;311(4):368–77.

19. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. 2013. pp. 814–22.

20. ICH harmonised tripartite guideline: statistical principles for clinical trials [Internet]. ich.org. 1998 [cited 2016 Oct 7]. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

21. Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. BMJ. 2013 Mar 21;346(mar21 1):f1135–5.

22. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. BMJ. 2009;338:b1732.

23. Anker SD, Agewall S, Borggrefe M, Calvert M, Jaime Caro J, Cowie MR, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J. 2014 Aug 7;35(30):2001–9.

24. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res. 2014 Oct;23(5):440–59.

25. Donaldson GW, Moinpour CM. Learning to live with missing quality-of-life data in advanced-stage disease trials. J Clin Oncol. 2005 Oct 20;23(30):7380–4.

26. Kyte D, Duffy H, Fletcher B, Gheorghe A, Mercieca-Bebber R, King M, et al. Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. PLoS ONE. 2014;9(10):e110229.

REFERENCES CHAPTER FIVE

1. Chalmers I, Glasziou P, Godlee F. All trials must be registered and the results published. BMJ. 2013 Jan 9;346:f105.

2. Golder S, Loke YK, Wright K, Norman G. Reporting of Adverse Events in Published and Unpublished Studies of Health Care Interventions: A Systematic Review. PLoS Medicine. 2016 Sep;13(9):e1002127.

3. Rennie D. Trial registration: a great idea switches from ignored to irresistible. JAMA. 2004 Sep 15;292(11):1359–62.

4. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. Lancet. 2004 Apr 24;363(9418):1341–5.

5. Jefferson T, Jones M, Doshi P, Spencer EA, Onakpoya I, Heneghan CJ. Oseltamivir for influenza in adults and children: systematic review of clinical study reports and summary of regulatory comments. BMJ. 2014 Apr 9;348:g2545.

6. Hartung DM, Zarin DA, Guise J-M, McDonagh M, Paynter R, Helfand M. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. Ann Intern Med. 2014 Apr 1;160(7):477–83.

7. Becker JE, Krumholz HM, Ben-Josef G, Ross JS. Reporting of results in ClinicalTrials.gov and high-impact journals. JAMA. 2014 Mar 12;311(10):1063–5.

8. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.Gov: a cross-sectional analysis. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM, editors. PLoS Medicine. 2009 Sep;6(9):e1000144.

9. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. BMJ. 2012 Jan 3;344(jan03 1):d7292–2.

10. Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. BMJ. 2016 Feb 17;352:i637.

11. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. BMJ. 2013 Oct 29;347(oct28 9):f6104–4.

12. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA. 2009 Sep 10;302(9):977–84.

13. COMPare Team. Tracking switched outcomes in clinical trials - Tracking switched outcomes in clinical trials [Internet]. compare-trials.org. 2016 [cited 2016 Nov 23]. Available from: http://compare-trials.org/

14. Clinical trial registration: transparency is the watchword. 2006 May 17;:1–3.

15. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet. 2004 Sep;364(9438):911–2.

16. Zarin DA, Tse T, Sheehan J. The proposed rule for U.S. clinical trial registration and results submission. The New England Journal of Medicine. 2015 Jan 8;372(2):174–80.

17. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010 Mar 23;340(mar23 1):c723–3.

18. van de Wetering FT, Scholten RJPM, Haring T, Clarke M, Hooft L. Trial Registration Numbers Are Underreported in Biomedical Publications. Scherer RW, editor. PLoS ONE. 2012 Nov 14;7(11):e49599.

19. Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg DAB. Issues in the registration of clinical trials. JAMA. 2007 May 16;297(19):2112–20.

20. Glanville JM, Duffy S, McCool R, Varley D. Searching ClinicalTrials.gov and the International Clinical Trials Registry Platform to inform systematic reviews: what are the optimal search approaches? J Med Libr Assoc. 2014 Jul;102(3):177–83.

21. HRA. Health Research Authority - Transparency, registration and publication - Health Research Authority [Internet]. hra.nhs.uk. [cited 2017 Aug 1]. Available from: http://www.hra.nhs.uk/resources/during-and-after-your-study/transparency-registration-and-publication/

22. Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med. 2011 Jul 5;155(1):39–51.

23. Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. BMJ. 2012;344:e813.

24. Chalmers I. Health Research Authority's great leap forward on UK trial registration. BMJ. 2013 Sep 25;347(sep25 2):f5776–6.

25. Krleza-Jeric K, Chan A-W, Dickersin K, Sim I, Grimshaw J, Gluud C. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). BMJ. 2005 Apr 23;330(7497):956–8.

26. Dechartres A, Ravaud P, Atal I, Riveros C, Boutron I. Association between trial registration and treatment effect estimates: a meta-epidemiological study. BMC Medicine. 2016;14(1):100.

27. Rasmussen N, Lee K, Bero L. Association of trial registration with the results and conclusions of published trials of new oncology drugs. Trials. 2009;10(1):116.

28. Emdin C, Odutayo A, Hsiao A, Shakir M, Hopewell S, Rahimi K, et al. Association of Cardiovascular Trial Registration With Positive Study Findings: Epidemiological Study of Randomized Trials (ESORT). JAMA Intern Med. 2014 Dec 29.

29. Wager E, Williams P, on behalf of the OPEN Project (Overcome failure to Publish nEgative fiNdings) Consortium. "Hardly worth the effort"? Medical journals' policies and their editors‘ and publishers’ views on trial registration and publication bias: quantitative and qualitative study. BMJ. 2013 Sep 6;347(sep06 1):f5248–8.

REFERENCES CHAPTER SIX

1. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010 Mar 23;340(mar23 1):c723–3.

2. Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet. 2005 Mar 19;365(9465):1159–62.

3. Baron G, Perrodeau E, Boutron I, Ravaud P. Reporting of analyses from randomized controlled trials with multiple arms: a systematic review. BMC Medicine. BMC Medicine; 2013 Mar 27;11(1):1–1.

4. Freidlin B, Korn EL, Gray R, Martin A. Multi-arm clinical trials of new agents: some design considerations. Clin Cancer Res. 2008 Jul 15;14(14):4368–71.

5. Parmar MKB, Carpenter J, Sydes MR. More multiarm randomised trials of superiority are needed. Lancet. 2014 Jul 26;384(9940):283–4.

6. Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. Lancet. 2005 May;365(9470):1591–5.

7. Wittes J. Sample size calculations for randomized controlled trials. Epidemiol Rev. 2002;24(1):39–53.

8. Proschan MA, Waclawiw MA. Practical guidelines for multiplicity adjustment in clinical trials. Control Clin Trials. 2000 Dec;21(6):527–39.

9. Boissel J-P. How to deal with multiple treatment or dose groups in randomized clinical trials? Another approach. Fundam Clin Pharmacol. 2007 Apr;21(2):155–6.

10. Hothorn LA. How to deal with multiple treatment or dose groups in randomized clinical trials? Fundam Clin Pharmacol. 2007 Apr;21(2):137–54.

11. EMA. Points to Consider on Multiplicity Issues in Clinical Trials [Internet]. ema.europa.eu. 2003 [cited 2016 Nov 23]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf

12. Wason JMS, Stecher L, Mander AP. Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? Trials. 2014 Sep 17;15:364.

13. Hennekens CHBJE, Mayrent SL. Epidemiology in medicine. Boston : Little Brown and Company, 1987; 1987.

14. Howard DR, Brown JM, Todd S, Gregory WM. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. Stat Methods Med Res. 2016 Sep 19.

15. Bender R, Lange S. Adjusting for multiple testing--when and how? Journal of Clinical Epidemiology. 2001 Apr;54(4):343–9.

16. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990 Jan;1(1):43–6.

17. Feise RJ. Do multiple outcome measures require p-value adjustment? BMC Med Res Methodol. 2002 Jun 17;2:8.

18. Bauer P, Chi G, Geller N, Gould AL, Jordan D, Mohanty S, et al. Industry, government, and academic panel discussion on multiple comparisons in a “real” phase three clinical trial. J Biopharm Stat. 2003 Nov;13(4):691–701.

19. Hopewell S, Collins GS, Boutron I, Yu L-M, Cook J, Shanyinde M, et al. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. BMJ. 2014 Jul 1;349(jul01 8):g4145–5.

REFERENCES CHAPTER SEVEN

1. ICH. General Considerations for Clinical Trials [Internet]. ich.org. 1997 [cited 2016 Nov 17]. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E8/Step4/E8_Guideline.pdf

2. Decullier E, Chan A-W, Chapuis F. Inadequate dissemination of phase I trials: a retrospective cohort study. PLoS Medicine. 2009 Feb 17;6(2):e1000034.

3. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. BMJ. 2013 Oct 29;347(oct28 9):f6104–4.

4. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. JAMA. 2014 Jan;311(4):368–77.

5. Kesselheim AS, Myers JA, Avorn J. Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. JAMA. 2011 Jun 8;305(22):2320–6.

6. Wang B, Kesselheim AS. Characteristics of efficacy evidence supporting approval of supplemental indications for prescription drugs in United States, 2005-14: systematic review. BMJ. 2015;351:h4679.

7. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. JAMA Intern Med. 2016 Oct 10.

8. Savulescu J. Two deaths and two lessons: is it time to review the structure and function of research ethics committees? J Med Ethics. 2002 Feb;28(1):1–2.

9. Savulescu J, Spriggs M. The hexamethonium asthma study and the death of a normal volunteer in research. J Med Ethics. 2002 Feb;28(1):3–4.

10. Savulescu J. Harm, ethics committees and the gene therapy death. J Med Ethics. 2001 Jun;27(3):148–50.

11. Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, et al. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. The New England Journal of Medicine. 2006 Sep 7;355(10):1018–28.

12. Duff G. Expert Scientific Group on Phase One Clinical Trials. 2006 Aug 4;:1–71.

13. St Clair EW. The calm after the cytokine storm: lessons from the TGN1412 trial. J Clin Invest. 2008 Apr;118(4):1344–7.

14. RSS. Report of the Working Party on Statistical Issues in First-in-Man studies [Internet]. rss.org.uk. 2007 [cited 2016 Nov 23]. Available from: http://www.rss.org.uk/Images/PDF/publications/rss-reports-statistical-issues-first-in-man-studies-2007.pdf

15. Eddleston M, Cohen AF, Webb DJ. Implications of the BIA-102474-101 study for review of first-into-human clinical trials. Br J Clin Pharmacol. 2016 Apr;81(4):582–6.

16. Callaway E, Butler D. Researchers question design of fatal French clinical trial. Nature. 2016 Jan 22.

17. Center for Drug Evaluation, Research. Therapeutic Biologic Applications (BLA) - Frequently Asked Questions About Therapeutic Biological Products. fdagov. Center for Drug Evaluation and Research.

18. European Medicines Agency. EU Clinical Trials Register – FAQs [Internet]. clinicaltrialsregister.eu. 2014 [cited 2016 Apr 20]. Available from: https://www.clinicaltrialsregister.eu/doc/EU_CTR_FAQ.pdf

19. Williams P, O'Connell D, Dewit O. First in Human Studies: Points to Consider in Study Placement, Design and Conduct [Internet]. abpi.org.uk. 2011 [cited 2016 Apr 22]. Available from: http://www.abpi.org.uk/our-work/library/guidelines/Documents/First%20in%20Human%20Studies.pdf

20. National Institute for Health Research. Clinical Trials Toolkit | Serious Adverse Event (SAE) or Serious Adverse Reaction (SAR) [Internet]. ct-toolkit.ac.uk. [cited 2016 Apr 22]. Available from: http://www.ct-toolkit.ac.uk/glossary/serious-adverse-event-sae-or-serious-adverse-reaction-sar

21. Orme M, Harry J, Routledge P, Hobson S. Healthy volunteer studies in Great Britain: the results of a survey into 12 months activity in this field. Br J Clin Pharmacol. 1989 Feb;27(2):125–33.

22. Lutfullin A, Kuhlmann J, Wensing G. Adverse events in volunteers participating in phase I clinical trials: a single-center five-year survey in 1,559 subjects. Int J Clin Pharmacol Ther. 2005 May;43(5):217–26.

23. Emanuel EJ, Bedarida G, Macci K, Gabler NB, Rid A, Wendler D. Quantifying the risks of non-oncology phase I research in healthy volunteers: meta-analysis of phase I studies. BMJ. 2015;350:h3271.

24. Dyer C. High Court rules that Health Research Authority acted unlawfully over trial transparency. BMJ (Clinical research ed). 2015;:h4194.

REFERENCES CHAPTER EIGHT

1. Rennie D. Trial registration: a great idea switches from ignored to irresistible. JAMA. 2004 Sep 15;292(11):1359–62.

2. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010 Mar 23;340(mar23 1):c723–3.

3. Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet. 2005 Mar 19;365(9465):1159–62.

4. Baron G, Perrodeau E, Boutron I, Ravaud P. Reporting of analyses from randomized controlled trials with multiple arms: a systematic review. BMC Medicine. BMC Medicine; 2013 Mar 27;11(1):1–1.

APPENDICES

APPENDIX 1: DEFINITION OF ITEMS IN SPIRIT CHECKLIST

	Variable	Options	Explanation
Ethics and General Protocol Characteristics			
1.	Country of Ethics Committee	Enter Text	
2.	Unique ID Number	Enter Text	UK is currently using ID###; Please consider a different version
3.	Name of Ethics Centre	Enter Text	
4.	Ethics Identification Number	Enter Text	
5.	Study Acronym	Enter Text	
6.	Study Title (Exact Quote)	Enter Text	
7.	Date of Approval by Ethics	Enter date	Enter as YYYY-MM-DD
8.	Clinical Area	Medical, Surgical, Paediatrics, Other	
9.	If medical area, choose from list	Neurology, cardiovascular, respiratory, gastro/intestinal, nephrology, rheumatology, infectious disease, oncology, intensive care, hematology, endocrinology, dermatology, anaesthetics, psychiatry, other	Best judgment – consider first: patient population, second: clinical outcome
10.	If pediatrics area, choose from list	Neurology, cardiovascular, respiratory, gastro/intestinal, nephrology, rheumatology, infectious disease, oncology, intensive care, hematology, endocrinology, dermatology,	Best judgment – consider first: patient population, second: clinical outcome

		anaesthetics, general surgery, neurosurgery, ophthalmology, ear-nose-throat (ENT), cardiothoracic, urology, orthopedics, plastic surgery, other	
11.	If surgical area, choose from list	General surgery, obstetrics / gynecology, neurosurgery, ophthalmology, ear-nose-throat (ENT), cardiothoracic, urology, orthopedics, plastic surgery, other	Best judgment – consider first: patient population, second: clinical outcome
12.	Trial Registry Name	Enter Text	Use acronym as needed (ISRCTN, EudraCT, ANZCTR, Clinicaltrials.gov) If not provided enter NR
13.	Trial Registration Number	Enter text	If not provided enter NR
Detailed Protocol Characteristics			
1.	Pilot Study	Yes, No	If pilot trial, please only extract protocol characteristics
2.	Dose finding trial	Yes, No	
3.	Hypothesis	Superiority, non-inferiority/equivalence, not labelled in this regard/unclear	Check all that apply
4.	Primary outcome(s) detail	Enter text	Copy and paste the primary outcome(s) from the protocol
5.	The setting for the majority of recruited patients is:	Community, outpatient clinic, emergency department, in-patients hospital care, intensive care unit, Unclear	Best judgment – try to avoid unclear Check all that apply
6.	Estimated sample size/number of participants	Enter number (total)	If not provided, enter "NR" If cluster randomized trial please provide total

			number of participants AND total number of clusters
7.	Number of Study Centres	Enter Text	If multicentre but number not provided, enter "multicentre" If not provided, enter "NR"
8.	If multicentre, national or multinational	National, International	
9.	Trial Design	Parallel, Crossover, Cluster, Factorial, Split Body, Other, Not Applicable	Check all that apply If non-randomized, enter "not applicable"
10.	Number of trial arms	Enter number	If unclear, enter "unclear"
11.	Name of intervention	Enter text	Copy and paste from protocol
12.	Intervention category	Drug Surgery/Invasive Procedure Device Biological/Vaccine Radiation Rehabilitation Behavioural/Lifestyle/Education/Counselling Dietary Supplement Other	
13.	Name of control	Enter text	Copy and paste from protocol
14.	Type of control	No Treatment/Standard Care, Active (drug/other treatment), placebo/sham	If Y is the experimental treatment on top of therapy X versus therapy X as the control please select "no treatment/standard care"; head –to-head comparisons would qualify for "active control"
15.	Name of funder(s)	Enter text	Copy and paste from protocol
16.	Funding	Government/public, Charity/private not	If document is written/sponsored by

		for profit, Solely industry/private for profit, Part industry/private for profit, not funded/only in house source, Not reported, Reported but type of funding unclear	industry/company (e.g. indicated by letter head), but funding in text not clearly stated, it should be labeled as industry/private for profit.
17.	Initiation/Sponsorship	Definitely industry initiated. Probably industry initiated, Probably investigator initiated, Definitely investigator initiated	Please use best judgement based on following criteria: - Initiator stated - Affiliation of protocol authors - Displayed company or institution logo - Included state-ments about data ownership or publication rights or statements about full funding by industry or public funding agencies
SPIRIT Checklist Items – Select Options Based on Whether Items are Reported			
1. Title	Basic study design	Yes, No	More details than randomised + trial phase necessary
2. Trial Registration – Registry	Trial identifier	Yes, No	
3. Protocol	Version Number	Yes, No	
4. Funding	Sources of financial and non-financial support	Yes, No	Should be explicitly stated as the funder
5. Roles and Responsibilities - Contributors	Names of protocol contributors	Yes, No	Should be stated explicitly that the persons were involved in preparing the protocol; list of principal investigators (PIs) or signature(s) of PIs on the protocol ok

6. Roles and Responsibilities – Sponsor Contact Information	Name of sponsor contact	Yes, No	The specific contact information for an individual should be provided
7. Roles and Responsibilities – Sponsor and Funder	Conduct	Yes, No	Roles and responsibilities of sponsor and funder described
8. Roles and Responsibilities – Steering Committee	General membership	Yes, No, Not Applicable	For each committee/group overseeing the trial, is the general membership and role described? Any level of detail is ok. If no steering committee, select "not applicable".
9. Background and rationale	Research question	Yes, No	
10. Background and rationale – choice of comparators	Comparator description	Yes, No	
11. Objectives	Objectives for each comparison	Yes, No	
12. Trial Design	Type of trial	Yes, No	
13. Study Setting	Countries where data will be collected	Yes, No	Specific country names required – continents not sufficient
14. Eligibility criteria	Inclusion and exclusion criteria for trial participants	Yes, No	
15.	Inclusion and exclusion criteria for study centres	Yes, No, Not Applicable	If single centre, select not applicable
16. Intervention (drug)	Generic name, Dose and Schedule	Yes, No, Not Applicable	All items must be reported. If non-drug intervention, select "not applicable"
17. Intervention (non-drug)	Setting	Yes, No, Not Applicable	If drug intervention, select "not applicable"
18.	Individuals administering interventions (e.g., expertise)	Yes, No, Not Applicable	If drug intervention, select "not applicable"
19. Interventions – Modifications	Standard criteria for modifications	Yes, No	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug

			dose change in response to harms, participant request, or improving/worsening disease)
20. Interventions – Adherence	Strategies to improve adherence or any procedures for monitoring adherence	Yes, No	
21. Interventions – Concommitaant care	Permitted care and interventions.	Yes, No	Relevant concomitant care and interventions that are permitted or prohibited during the trial
22. Outcomes – Primary Outcome only	Specific measurement variable	Yes, No, Not Applicable	If more than one primary outcome, select the outcome that is used for the sample size calculation. If more than one sample size calculation, randomly select from the primary outcome. If no primary outcome, select not applicable.
23.	Analysis metric	Yes, No, Not Applicable	E.g., Change from baseline, Final value, Time-to-event
24.	Time point	Yes, No, Not Applicable	No time point needed for PFS, ORR. If overall survival and risk ratio, then they need to give a time point. If it's time to event, then would not expect time point.
25. Participant timeline	Timing of visit	Yes, No	
26. Sample size	Sample size (total and per group)	Yes, No	
27.	Outcome used	Yes, No, Not Applicable	Whether justification/rationale provided. If not derived statistically, select not applicable
28.	Assumed values for outcome in each study group	Yes, No, Not Applicable	Whether justification/rationale provided. If not derived

			statistically, select not applicable If the minimally important difference, the value for one group should be given as well.
29.	Rationale or reference for assumed outcome values	Yes, No, Not Applicable	Whether justification/rationale provided. If not derived statistically, select not applicable If no for above, then not applicable for this one
30.	Statistical test	Yes, No, Not Applicable	Whether justification/rationale provided. If not derived statistically, select not applicable
31.	Alpha	Yes, No, Not Applicable	Whether justification/rationale provided. If not derived statistically, select not applicable
32.	Power	Yes, No, Not Applicable	Whether justification/rationale provided. If not derived statistically, select not applicable
33.	Adjustment for missing data, if relevant	Yes, No, Not Applicable	Whether justification/rationale provided. If not derived statistically, select not applicable
34.	Rationale for intended sample size if not derived statistically	Yes, No, Not Applicable	Whether justification/rationale provided. If derived statistically, select not applicable
35. Recruitment	Location of participant recruitment	Yes, No	Reporting question! E.g. emergency department, community
36.	Who will recruit participants	Yes, No	E.g. surgeon, study nurse
37.	Expected recruitment rates	Yes, No	
38.	Estimate for eligible patients	Enter number	If not provided, enter "NR"

39.	What is the estimated duration of the recruitment period	Enter number and unit (e.g. 1.5 years)	If not provided, enter "NR"
40.	Plans to monitor recruitment during trial	Yes, No	Any level of detail ok (e.g. DSMB monitoring recruitment).
41.	Financial and non-financial incentives for participants	Yes, No	Any statement referring to incentives (including any reimbursements) for participants ok
42.	Financial and non-financial incentives for investigators	Yes, No	Any statement referring to incentives for investigators ok
43. Allocation – Sequence Generation	Random allocation - Method of sequence generation	Yes, No, Not Applicable	If non-random allocation or single arm, select "not applicable". Mentioning of statistician only is not sufficient.
44.	Random allocation - Allocation ratio	Yes, No, Not Applicable	If non-random allocation or single arm, select "not applicable".
45.	Random allocation - Type of randomization	Yes, No, Not Applicable	If non-random allocation or single arm, select "not applicable". If obvious from the protocol that it is "simple" randomisation, check "yes" even if not explicit
46.	Non-random allocation - Method	Yes, No, Not Applicable	If random allocation, select "not applicable".
47.	Non-random allocation - Rationale	Yes, No, Not Applicable	If random allocation, select "not applicable".
48. Allocation – concealment mechanism	Allocation concealment mechanism	Yes, No, Not Applicable	If non-random allocation or single arm, select "not applicable".
49. Allocation - Implementation	Who will enrol/assign participants	Yes, No, Not Applicable	If non-random allocation or single arm or cluster-randomized, select "not applicable".
50. Blinding/Masking	Blinding status of participants	Yes, No, Not Applicable	If unblinded, select "not applicable". If open label but not specifically stated as unblinded, select "No".

51.	Blinding status of care providers	Yes, No, Not Applicable	If unblinded, select "not applicable". If open label but not specifically stated as unblinded, select "No".
52.	Blinding status of outcome assessors	Yes, No, Not Applicable	If unblinded, select "not applicable". If open label but not specifically stated as unblinded, select "No".
53. Blinding – Emergency Unblinding	Circumstances/conditions when unblinding is permissible	Yes, No, Not Applicable	If unblinded, select "not applicable".
54. Data Collection Methods	Data collection personnel	Yes, No	Should be mentioned who will collect outcome data (e.g. study nurse, treating physician)
55. Data Collection Methods - Retention	Strategies to promote participant retention and complete follow-up	Yes, No	Any level of detail ok.
56. Data Management	Data entry and coding processes	Yes, No	Plans for data entry, coding, security, and storage, including any related processes to promote data quality
57. Statistical Methods - Outcomes	Main analysis for primary outcome, including analysis methods for statistical comparisons	Yes, No	
58.	How missing data will be handled	Yes, No	
59.	Effect measure	Yes, No	If they give the anova/t-test and list the variable, then we know it's a difference in means so give a "Yes"
60.	Significance level	Yes, No	
61.	Intended use of confidence intervals	Yes, No	
62. Statistical Methods – Additional Analyses	Definition of subgroup categories	Yes, No, Not Applicable	If no subgroup analysis, select "Not Applicable"
63. Statistical Methods – Analysis	Which participants will be included in the main analyses	Yes, No	

population and missing data			
64. Data Monitoring Committee	DMC included? If no, reason given?	Yes, No	
65. Data Monitoring – Interim Analyses	Planned number of interim analyses	Enter number	If no number provided, enter "NR"
66.	Purpose of interim analyses	Benefit Harm Futility Sample size recalculation Other No reason provided Not applicable	Select all that apply
67.	If other please specify	Enter text	
68.	Who has ultimate authority to stop the trial	Yes, No, Not Applicable	If no interim analysis, select "not applicable".
69. Harms	Anticipated/Unanticipated	Yes, No	
70. Auditing	Procedures and anticipated frequency of audits	Yes, No, Not Applicable	Auditing involves periodic independent review of core trial processes and documents. It is distinct from routine day-to-day measures to promote data quality. If protocol states no audit planned, select "not applicable".
71. Research Ethics Approval	Where approval has been obtained, or plans for seeking approval	Yes, No	Should be always "Yes" in this study
72. Protocol Amendments	Process for making amendments	Yes, No	
73. Consent or Assent	Consent process	Yes, No	
74. Consent or Assent – Ancillary Studies	Process to obtain additional consent for collection and use of data and biological specimens	Yes, No, Not Applicable	If no ancillary studies mentioned, select "not applicable".
75. Confidentiality	How data will be collected, kept secure, and maintained during the trial	Yes, No	

76. Declaration of Interests	Financial and other competing interests	Yes, No	
77. Access to Data	Who will have access to full dataset after trial completion	Yes, No	
78. Ancillary and post-trial care	Any plans to provide or pay for ancillary care during the trial	Yes, No	
79. Dissemination Policy - Results	Plans to disseminate trial results to key stakeholders/publication	Yes, No	Any explicit plans justify a "yes"
80. Dissemination Policy - Authorship	Authorship eligibility criteria	Yes, No	Any explicit criteria or reference to criteria (e.g. of a journal) or section on authorship ok.
81. Dissemination Policy – Reproducible Research	Full trial protocol	Yes, No	
82. Informed Consent Materials	Model consent and/or assent forms	Yes, No	Actual forms should be provided.
83. Biological Specimens	Details of specimen collection	Yes, No, Not Applicable	If no biological specimens mentioned, select "not applicable".

APPENDIX 2: EXAMPLES OF PATIENT REPORTED OUTCOME MEASURES

Patient Reported Outcome Measures Reported in Clinical Trials Protocols in the ASPIRE Cohort
Patient comfort during gastroscopy, as measured by the Global Rating Scale scoring system for patient comfort. Carer self-reported depression and anxiety using the Hospital Anxiety and Depression Scale. The QUALEFFO 41 is a disease specific measure of health related quality of life (Qol) applicable to patients with established vertebral osteoporosis, independently of whether or not they have back pain and is designed to be used as an evaluative instrument in clinical trials. It is a self-administered questionnaire that provides scores on five domains; pain, physical function, social function, general health perception, mental performance, and a total score. All scores are expressed as a normalised score from 0-100, where higher scores represent worse Qol. It is validated and reliable and has been shown to be responsive in clinical trials of physiotherapy treatment The key outcome measures will be levels of insomnia as assessed by the Insomnia Severity Index (ISI), and levels of delusions and hallucinations as assessed by the Psychotic Symptoms Rating Scale (PSYRATS). Recurrence rates of symptoms of haemorrhoids at 1 year follow up Post-operative pain scores and analgesia requirements Progression□free survival, defined as the time interval between randomisation and disease progression or death
Outcomes Not Classified as Patient Reported Outcome Measures
Progression-free survival, defined as the time interval between randomisation and disease progression or death Exercise capacity at 6 weeks as assessed by the 6-minute walk test distance in metres. The primary outcome of activity limitation will be measured using the Overall neuropathy limitations scale (ONLS)2. The ONLS is a 12 item ordinal scale of activity limitation developed from the ODSS which includes upper and lower limb sub-scales, validated for use in people with IN and for interview or observed administration.

APPENDIX 3: ORIGINAL RESEARCH PUBLICATIONS CONTRIBUTING TO THIS THESIS

Ayodele Odutayo, Bethan Copsey, Susan Dutton, Jonathan Cook, Sally Hopewell, Douglas Altman. Assessment of the Design, Conduct and Dissemination of Phase I Trials. *JAMA*. 2017. 317(17):1799-1801

APPENDIX 4: OTHER NOTABLE PUBLICATIONS DURING THIS DPHIL

Ayodele Odutayo, Peter Gill, Shaun Shepherd, Aquila Akingbade, Sally Hopewell, Karthik Tennankore, Benjamin Hunn, Connor Emdin. Income Disparities in Absolute Cardiovascular Risk and Cardiovascular Risk Factors in the United States, 1999-2014. *JAMA Cardiology* 2017. doi:10.1001/jamacardio.2017.1658.

Ayodele Odutayo, Connor Emdin, Allan Hsiao, Mubeen Shakir, Bethan Copsey, Susan Dutton, Virginia Chiocchia, Michael Schlussel, Peter Dutton, Corran Roberts, Douglas G Altman, Sally Hopewell. Association between trial registration and positive study findings: cross sectional study (Epidemiological Study of Randomized Trials-ESORT). *British Medical Journal*. 2017;356:j917.

Ayodele Odutayo, Christopher Wong, Allan Hsiao, Sally Hopewell, Douglas G Altman and Connor Emdin. Atrial fibrillation and the risks of cardiovascular disease, renal disease and death: a meta-analysis. *British Medical Journal*. (2016). 354:i4482.

Ayodele Odutayo, Christopher Wong, Michael Farkouh, Douglas G Altman, Sally Hopewell, Connor Emdin and Benjamin Hunn. Acute Kidney Injury and Long-Term Risk for Cardiovascular Events and Mortality. *Journal of the American Society of Nephrology*. (2016). June 13.

Ayodele Odutayo, Kazem Rahimi, Allan Hsiao and Connor Emdin. Blood Pressure Targets and Absolute Cardiovascular Risk. *Hypertension*. (2015). Vol. 66 (2) pp 280-285.

Connor Emdin*, **Ayodele Odutayo*,** Allan Hsiao, Mubeen Shakir, Kazem Rahimi, Sally Hopewell and Douglas Altman. The association between randomized trial evidence and global burden of disease: an observational study. *British Medical Journal.* (2015). 350:h117. *contributed equally

Connor Emdin*, **Ayodele Odutayo*,** Allan Hsiao, Mubeen Shakir, Kazem Rahimi, Sally Hopewell and Douglas Altman. Registration of cardiovascular trials and its association with positive study findings. *JAMA Internal Medicine.* (2015) vol 175 (2) pp 304-307. *contributed equally.

APPENDIX 5: CONFERENCE ABSTRACTS AND PRESENTATIONS DURING THIS DPHIL

Evangelos Oikonomou, **Ayodele Odutayo,** Ioannis Akoumianakis, Alexios Antonopoulos, Sheena Thomas, Laura Herdman, Marialena Trivella, Gary Collins, Mario Petrou, Charlambos Antionades. Identifying Biological Predictors of Early Saphenous Vein Graft Failure: A Meta-Analysis of 5134 Patients with Angiographic Follow Up in SAFINOUS-CABG Collaborative Group. Oral Presentation at the American Heart Association Conferece, November 2016.

Ayodele Odutayo, Bethan Copsey, Angela MacCarthy, An-Wen Chan, Douglas Altman and Sally Hopewell. Registration of Clinical Trials in the United Kingdom. Oral Presentation at the REWARD|EQUATOR Conference, September 2015.

Ayodele Odutayo, Bethan Copsey, Vipul Jairath, An-Wen Chan, Sally Hopewell and Douglas Altman. Assessing the Use of Patient Reported Outcomes in Protocols for Randomized Controlled Trials Approved in the United Kingdom in 2012. Oral Presentation at the REWARD|EQUATOR Conference, September 2015.

Ayodele Odutayo, Arti Bhasin and Michelle Hladunewich. Pregnancy Outcomes in Women with Primary Glomerular-Based Disease: A Systematic Review. Poster at the 2014 American Society of Nephrology Meeting, November 2014.

Ayodele Odutayo, Connor Emdin, Allan Hsiao, Kazem Rahimi, Sally Hopewell and Douglas Altman. Registration of Randomized Controlled Trials Published in PubMed Indexed Journals: A Cross-sectional Analysis. Oral at the 22nd Annual Cochrane Colloquium, September 2014.

Connor Emdin, **Ayodele Odutayo,** Allan Hsiao, Kazem Rahimi, Sally Hopewell and Douglas Altman. Relationship Between Global Burden of Disease and Quantity of Published Randomized Controlled Trials. Oral at the 22nd Annual Cochrane Colloquium, September 2014.

APPENDIX 6: PEER REVIEW RESPONSIBILITIES DURING THIS DPHIL

Journals (PubMed Indexed Only)

Journal of the American College of Cardiology (2017)

Nephrology Dialysis Transplantation (2017)

Clinical Interventions in Aging (2 manuscripts, 2016)

Journal of Evaluation in Clinical Practice (2016)

Cochrane Heart Group (2 manuscripts; 2016)

Journal of Cardiac Failure (2016)

Drug, Healthcare and Patient Safety (2016)

Journal of Clinical Epidemiology (2016)

Trials (2 manuscripts; 2015)

Drug Design, Development and Therapy (2015)

Granting Agencies

European Research Council (2015 and 2016)

APPENDIX 7: COURSES COMPLETED DURING THIS DPHIL

Stata Autumn School (Missing Data, Multilevel Modelling, Survey Data and Network Meta-analysis), Timberlake Consultants Training, London, United Kingdom (4 days, September 2015)

R: An Introduction, University of Oxford, Oxford, United Kingdom (2 days, April 2015)

R Programming, Online Course, Coursera (2 weeks, January 2015)

Designing Clinical Research and Biostatistics, University of Oxford, Oxford, United Kingdom (2 days, November 2014)

Designing Randomized Controlled Trials, University of Oxford, Oxford, United Kingdom (5 days, September 2014)