

Items for consideration in a reporting guideline for mediation analyses: a Delphi study

Aidan G Cashin ^{a,b}	Email: a.cashin@neura.edu.au
James H McAuley ^{a,c}	Email: j.mcauley@neura.edu.au
Sarah E Lamb ^{d,e}	Email: sarah.lamb@ndorms.ox.ac.uk
Sally Hopewell ^d	Email: sally.hopewell@csm.ox.ac.uk
Steven J Kamper ^{f,g}	Email: steven.kamper@sydney.edu.au
Christopher M Williams ^{g,h}	Email: christopher.m.williams@hnehealth.nsw.gov.au
Nicholas Henschke ⁱ	Email: nicholas.henschke@sydney.edu.au
Hopin Lee ^{d,h*}	Email: hopin.lee@ndorms.ox.ac.uk

^aNeuroscience Research Australia, Sydney, Australia

^bPrince of Wales Clinical School, Faculty of Medicine, University of New South Wales, Sydney, Australia

^cSchool of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, Australia

^dOxford Clinical Trials Research Unit and Centre for Statistics in Medicine, Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK

^eInstitute of Health Research, University of Exeter Medical School, Exeter, UK

^fInstitute for Musculoskeletal Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

^gCentre for Pain, Health and Lifestyle, Newcastle, Australia

^hSchool of Medicine and Public Health, University of Newcastle, Newcastle, Australia

ⁱSchool of Public Health, University of Sydney, Sydney, Australia

*Corresponding Author:

Dr Hopin Lee

Botnar Research Centre

Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Headington, Oxford, UK OX3 7LD

Tel: +44 018 652 26493

Email: hopin.lee@ndorms.ox.ac.uk

Pages: 18

Word count: 2574 (up to 3500)

Figures and tables: 2 figure, 2 tables

SUMMARY BOX

What is already known about this subject

- Mediation analysis is a widely used quantitative method for investigating the mechanisms of interventions and exposures in randomised controlled trials and observational studies.
- Reporting of studies that use mediation analysis is often heterogeneous and incomplete.

What are the new findings?

- Nineteen international experts reached consensus on 34 items that should be reported in studies that use mediation analysis. These experts also contributed 60 qualitative comments to help refine and prioritise items for the final reporting guideline.

How might it impact on research practice in the foreseeable future?

- The level of agreement and importance of these 34 items, in addition to the experts' comments will inform a consensus meeting to confirm the final list of reporting items that will be included in a reporting guideline for studies using mediation analysis.
- Complete and transparent reporting of studies that use mediation analysis can facilitate clinical application, reproducibility, and evidence synthesis.

ABSTRACT

Objective

Mediation analysis is a widely used quantitative method for investigating how interventions and exposures in randomised controlled trials and observational studies have an effect on healthcare outcomes. This study aimed to assess the importance of items that should be considered in a consensus meeting aimed at developing a guideline for reporting mediation analyses.

Study Design and Setting

A Delphi panel of international experts were asked to rate the importance of a list of items for inclusion in a guideline for reporting mediation analyses. Thresholds for disagreement and consensus on importance for inclusion were specified *a priori*. We used the Research AND Development/University of California Los Angeles appropriateness method to quantitatively assess the importance for inclusion and panel agreement.

Results

Nineteen expert panellists (10 female) from seven countries agreed to participate. All panellists contributed to all three rounds conducted between 10/06/2019 and 06/11/2019. The panel reached consensus on 34 unique reporting items for study design, analytic procedures and effect estimates, with 3 items rated 'optional'. Panellists added one extra item and provided 60 qualitative comments for item refinement and prioritisation.

Conclusion

This Delphi study used a rigorous consensus process to reach consensus on 34 reporting items for studies that use mediation analysis. These results will inform a consensus meeting that will consolidate a core set of recommended items for reporting mediation analyses.

Key words

Mediation analysis, mechanisms, reporting guideline, Delphi

INTRODUCTION

Health exposures and interventions have their effects on outcomes through causal mechanisms.[1] Mediation analyses of randomised controlled trials and observational studies have been used to understand the mechanisms by which an intervention or exposure has an effect on a given outcome.[2,3] Mediation analysis is an analytical method that is used to separate the effect of an exposure or intervention into an ‘indirect effect’ which works through the mechanism(s) of interest, and a ‘direct effect’ which works through all other mechanisms (Figure 1).[4] The results from mediation analyses are used to develop and optimise interventions, and to inform the implementation of policy and clinical care.[5–7] Research into understanding causal mechanisms has been endorsed by the UK Medical Research Council (MRC)[1] and National Institute for Health Research (NIHR) through their Efficacy and Mechanism Evaluation Programme[8] and is also recommended by other national funding bodies.

Although the application of mediation analysis to randomised trials and observational studies is increasing exponentially,[9] recent systematic reviews have identified inadequate reporting of mediation analyses.[10–17] Vo et al (2020) found that 96% of randomised controlled trials that use mediation did not report a sample size calculation and 70% did not report any underlying assumptions of mediation analyses.[17] There are no specific reporting guidelines for studies that use mediation analysis; and generic guidelines such as the Consolidated Standards of Reporting Trials (CONSORT) [18] and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)[19] do not cover the additional aspects of study design, analysis, and effects that should be reported in a mediation analysis. Numerous papers have highlighted the need for a specific reporting guideline for mediation analyses.[10,15–17]

In 2017 we formed an international working group to develop A Guideline for Reporting Mediation Analysis (AGReMA).[20] The AGReMA project methods are in accordance with the Guidance for Developers of Health Research Reporting Guidelines by Moher et al (2010).[21] After the completion of an overview of systematic reviews to assess the extent of reporting inadequacies,[16] and reviewing specific aspects of reporting quality,[10,15,17] we conducted a Delphi study to reach expert consensus on items that should be considered in a consensus meeting that will consolidate a core set of recommended items for reporting mediation analyses.

METHODS

A Delphi study is a structured method for achieving consensus among a panel of experts on a given question or topic.[22] The process involved developing an initial exhaustive list of reporting items from previous reviews and consultation with experts, followed by a series of online surveys to: (1)

assess the level of agreement on the initial list of reporting items; (2) elicit additional items and clarification on items to refine the initial list; and (3) identify which items are considered critically important in reporting mediation studies. We preregistered the protocol prior to data collection on 21 May 2019 on the *Open Science Framework*, (<https://osf.io/xwka6>). The University of New South Wales Human Research Ethics Advisory Panel granted ethical approval (HC16599).

Selection of preliminary items

We generated an exhaustive preliminary list of 33 items to be considered in the first round of the Delphi based on a scoping review of existing methods and reporting guidance documents for mediation analyses, and from the findings of our recent overview of systematic reviews.[16] In addition, we examined items from the CONSORT and STROBE checklist [18,19] to identify potential reporting items that could be adapted for mediation studies.

Selection of experts

We compiled a list of experts who had experience in the development and application of mediation analysis. Experts included those who had published original research papers involving mediation analysis or systematic reviews of mediation studies; or a methodological/statistical paper on mediation analysis; or a textbook on mediation analysis. Experts were identified through a variety of sources including an overview[16] and a scoping review of the literature, and through informal discussions with key stakeholders. Recruitment was an iterative process, with the final list of experts for the Delphi panel determined by the AGRReMA working group. The panel size for the Delphi consensus methodology[23] is best qualified in terms of its ‘expertness’ rather than its size,[24] differing to surveys which require larger sample sizes for statistical power. We, like most other Delphi studies [25,26], used the definition by Fink et al. (1984) that an expert “should qualify for selection because they are representative of their profession, have power to implement the findings, or because they are not likely to be challenged as experts in the field”. [27]

Delphi procedure

The Delphi procedure involves a series of ‘rounds’, where panellists independently and anonymously contribute and rank items until pre-defined consensus is reached. A central component of the Delphi procedure is the provision of summary feedback to panellists after each round to encourage the reflection on judgements for subsequent rounds to assist transforming individual opinion into group consensus.[22] The Delphi technique enables the anonymous contribution of experts across geographic locations while reducing the limitations imposed by group dynamics such as the dominance of a few panel members.[22,28,29] We collaborated with CLINVIVO (www.clinvivo.com), an independent company that designs and facilitates web-based Delphi studies.

CLINVIVO developed a bespoke electronic data capture program and co-ordinated the data collection and analysis for each Delphi round.

The first Delphi round started on 10/06/2019 and the third (final) round concluded on 06/11/2019. Each round took 3 to 5 weeks. Panellists were sent three reminders including one personalised reminder from the authorship team to ensure an adequate response rate.

Round one

All panellists were sent a link that included a description of the project, demographic questionnaire, and the preliminary list of reporting items for consideration. Panellists rated the importance of each item on a 9-point Likert scale (1, 'not important', to 9, 'critically important') and scored their confidence in their ratings (1, 'not confident', to 9, 'very confident'). Panellists also provided unlimited free-text suggestions on item wording, where applicable, and proposed additional items for consideration in the subsequent round.

Round two

The link to round two included a summary of results from round one (mean scores and standard deviations, median scores and inter-percentile ranges (IPR), histograms and descriptive labels of importance and agreement level, anonymised qualitative comments), together with the panellists' own score of importance for each item.[30] Newly nominated items and suggested re-wording of items from round one were also presented. Panellists re-rated the importance of each item in the light of the round one results and re-scored their confidence in their ratings on a 9-point Likert Scale (1, 'not confident', to 9, 'very confident'). Panellists were informed that items scored for importance ≤ 3 would be excluded, and items scored ≥ 7 would be considered critically important for the draft guideline to be discussed at the consensus meeting.

Round three

The link to round three included a summary of results from round two (mean/medians and standard deviations/IPRs, histograms, descriptive labels of importance and agreement level, anonymised qualitative comments), together with the panellist's own scores on importance. Panellists were informed which items had reached a median importance score ≥ 7 to be included and which items reached a median score ≤ 3 to be excluded. Panellists categorised the remaining items for which consensus had not been reached (median score 4-6 or where disagreement exists – see analysis below) as either: 1) 'Include as optional' or 2) 'Exclude'. Panellist also scored their confidence in their ratings on a 9-point Likert Scale (1, 'not confident', to 9, 'very confident').

Data analysis

We used descriptive statistics to summarise demographic data from included panellists. We coded and thematically grouped the free text comments from round one and two to identify the key issues and common themes. Newly nominated items were discussed and included in the subsequent round if there were considered substantially different to the current items, if not, they were considered in the rewording of similar items. We used the Research AND Development/University of California Los Angeles (RAND/UCLA) appropriateness method to analyse the scores from each round. We modified this approach by asking panellists to rate ‘importance’ rather than ‘appropriateness’. The RAND/UCLA appropriateness method considers the median panel rating and dispersion of each panel rating to provide an index of importance and agreement.[23] This involves calculating the median score, the IPR (30th and 70th) and the inter-percentile range adjusted for symmetry (IPRAS) for each item being rated. We considered agreement to be present when the IPR was equal to or less than the IPRAS and disagreement to be present when the IPR was greater than the IPRAS.[23] For the analysis of the round one and two, we considered the RAND/UCLA definitions for consensus on items to be considered for the reporting guideline as [23]:

- ‘Include’: panel median of 7-9 for importance, without disagreement
- ‘Uncertain’: panel median of 4-6 for importance, or any median with disagreement
- ‘Exclude’: panel median of 1-3 for importance, without disagreement

For the analysis of round three, we considered consensus for items to be considered for the reporting guideline as [25]:

- ‘Include as optional’: panel majority as include
- ‘Exclude’: panel majority as exclude

RESULTS

Demographics of participants

Of the 40 invited experts, 19 consented and agreed to participate. Table 1 describes the panel demographics.

Table 1. Demographics of experts

Characteristic	Number of Experts (n=19)
Age, mean (SD)	45.5 (8.7)
Female, number (%)	10 (53%)
Doctorate (PhD) qualification, number (%)	19 (100%)
<i>Country</i> , number (%)	
USA	8 (42%)
England/UK	3 (16%)

Netherlands	3 (16%)
Australia	2 (11%)
Canada	1 (5%)
Denmark	1 (5%)
Belgium	1 (5%)
<i>Academic title, number (%)</i>	
Assistant Professor	8 (42%)
Professor	7 (37%)
Researcher	3 (16%)
Associate Professor	1 (5%)
<i>Primary job title, number (%)</i>	
Statistician	6 (32%)
Epidemiologist	5 (26%)
Clinical researcher	2 (11%)
Health services researcher	1 (5%)
Other	5 (26%)
<i>Role in development and/or application of mediation analysis, number (%)</i>	
Application (applied mediation analysis)	16 (84%) [#]
Synthesise (review and summarise studies that apply mediation analysis)	15 (79%) [#]
Methodology (developed study designs)	10 (53%) [#]
Statistics (developed statistical methods)	9 (47%) [#]
Other	1 (5%) [#]

[#] Counts and percentages are not cumulative, panellists were able to select more than one option

Round one

In round one, the 19 panellists rated 33 items of which 30 achieved consensus (Figure 2) as ‘important items’ (median score ≥ 7), three achieved consensus of ‘uncertain importance’ (median score 4-6) and zero were considered ‘not important’ items (median score ≤ 3) (Table 2). The IPRAS was greater than the IPR for all item scores indicating no disagreement between experts in round one. The panellists provided 37 qualitative comments suggesting clarifications for 10 items and the addition of one item.

Table 2 here

Round two

All 19 panellists completed round two. Panellists rated a total of 14 items, including 10 items adapted from round one to improve clarity (items 1, 8, 7, 9, 12, 14, 18, 19, 24, 26, 31 in Table 2), three items which were rated of ‘uncertain importance’ (items 7, 17, 30 in Table 2) and one newly suggested item

(item 34 in Table 2). The 20 items that reached consensus in round one were not re-rated. Of the 14 items rated in round two, 11 achieved consensus as ‘important’ (median score ≥ 7) (items 1, 8, 9, 12, 14, 18, 19, 24, 26, 31, 34 in Table 2), three remained ‘uncertain’ (median score 4-6) (items 7, 17, 30 in Table 2) and zero were considered ‘not important’ items (median score ≤ 3). The IPRAS was greater than the IPR for all item scores indicating no disagreement between experts in round two. The panellists provided 14 qualitative comments on 9 items. These comments provided feedback on the revised wording, the additional item from round one, and the three ‘uncertain’ items. The experts suggested that the three ‘uncertain’ items (items 7, 17, 30, Table 2) were not specific or relevant to mediation studies and should be considered ‘optional’ items.

Round three

All 19 panellists completed round three. Panellists reached consensus to include the three remaining items (items 7, 17, 30) as ‘optional’ items. The panellists provided 9 additional qualitative comments supporting the ‘optional’ consideration for the three items (items 7, 17, 30).

In summary, round one reached consensus to include 20 items, round two reached consensus to include an additional 11 items, and round three reached consensus to include the three final items as ‘optional’ (Figure 2). No item from the preliminary list reached consensus for exclusion. The final list included 34 items which achieved consensus to be considered for AGReMA.

Summary of free-text response

The panellists provided 60 item-specific and general comments during the three Delphi rounds. The comments helped revise 13 (39%) of the preliminary Delphi items, and introduced one additional item. Panellists comments also raised concern on the number of items and the degree of detail; “some of these questions concern overlapping topics, so I would simplify them”. Further, some panellists described the need for the guideline to be more specific to mediation; “I think some of your entries are not specific to mediation analysis but general for any scientific papers”; and the relevance of the guideline for all mediation publication types; “I think [the] reporting guideline also need to take into account whether the entire paper is on mediation, or whether this is just a secondary analysis with the primary analysis (in the same paper) being the analysis of total effects”. All item-specific and general comments were circulated to panellists in each subsequent round.

DISCUSSION

A group of international experts in mediation analysis reached consensus on a preliminary list of 34 items that should be considered in the next stage of the AGReMA project, a consensus and guideline development meeting. The Delphi method is an important step that engages key stakeholders in the

process of developing a reporting guideline.[21] We included a representative panel of experts who achieved a high level of agreement ($\geq 7/9$) and confidence ($\geq 7/9$) in item ratings. In each round of the Delphi, there was no evidence of disagreement between panellists based on the RAND/UCLA appropriateness method. This evidence generated from a group of international experts contributes to the growing literature of reporting practices in mediation analysis and will guide the development of AGReMA.

The panellists' item-specific and general comments provided valuable insights for item refinement and prioritisation. The comments also raised concerns to be further addressed at the consensus meeting, including the need for a concise, mediation specific guideline that is suitable for all mediation analysis publication types. The recent Guidance for Reporting Involvement of Patients and the Public (GRIPP2) reporting guideline produced two versions of their reporting checklist: a long form for studies where patient and public involvement is the primary focus, and a short form for studies where it is a secondary or tertiary focus.[31] This format could also be adopted in AGReMA as mediation analyses of randomised trials and observational studies are also reported as primary and secondary publications.

Although the preliminary list contained an exhaustive list of reporting items, all items eventually reached consensus for inclusion. The majority of items were related to study methods (25/34, 74%). This is in keeping with the function of reporting guidelines to describe study features that influence the risk of bias.[32] Our current list of 34 items is larger than CONSORT (25 items) and STROBE (22 items). In this Delphi, we purposefully split broad concepts into separate items. In the consensus meeting we may merge related items to reduce the number of items to enhance useability.

Strengths and limitations

We prospectively registered the protocol prior to data collection and used robust methodology to conduct this Delphi study. We used *a priori* thresholds of disagreement and consensus. We used the RAND/UCLA method to measure panel disagreement and to establish clear boundaries of 'importance' for inclusion. The Delphi rounds and data collection were externally conducted by CLINVIVO to avoid influence from the AGReMA working group.[33] Finally, we obtained 100% follow-up and completion rates at all rounds. This in part was influenced by timely and personalised reminders distributed through CLINVIVO's bespoke electronic data capture program and by recruiting a clearly defined expert sample of engaged panellists.

There are some limitations to our study. The results from this study are only as externally valid as the representation of the expert panel. We attempted to include a diverse, international panel of experts who represented key stakeholders. Nineteen (49%) of the 40 invited experts agreed to participate in

this study which is consistent with other Delphi recruitment rates.[26,34] However, we cannot be sure that our sample accurately represents the experts who chose not to participate. The definition of an expert itself has been debated. To safeguard the validity of our findings, we used systematic methodology to identify experts who meet the definition suggested by Fink et al. (1984).[27] Finally, we followed an iterative process of re-rating uncertain items for inclusion/exclusion to facilitate consensus between panellists, however, we did not provide an opportunity for panellists to re-rate the 20 items that reached consensus for inclusion following round one.

CONCLUSION

Based on a Delphi process, a preliminary list of 34 reporting items for studies that use mediation analysis were identified. These findings will inform a consensus meeting to develop a guideline for reporting mediation studies that is relevant across disciplines, study designs and publication types.

DECLARATIONS

List of abbreviations

None

Ethics approval and consent to participate

Ethics approval was obtained from the University of New South Wales Human Research Ethics Advisory Panel D: Biomedical, approval number HC16599.

Submission declaration and verification

This article has not been previously published

Availability of data and materials

The dataset used and analysed during this study is available from the corresponding author on reasonable request.

Funding

This work was supported by project funding from the Berkeley Initiative for Transparency in the Social Sciences, a program of the Center for Effective Global Action (CEGA), with support from the Laura and John Arnold Foundation. The funding body did not contribute to the design of the study, the collection, analysis and interpretation of data and in writing the manuscript.

Competing interests

AC is supported by the University of New South Wales Prince of Wales Clinical School Postgraduate Research Scholarship and a NeuRA PhD Candidature Supplementary Scholarship, and is a Catalyst for the Berkeley Initiative for Transparency in the Social Sciences. HL is funded by the National Health and Medical Research Council (APP1126767); National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust; received project funding from the Berkeley Initiative for Transparency in the Social Sciences, a program of the Center for Effective Global Action (CEGA), with support from the Laura and John Arnold Foundation; and is a Catalyst for the Berkeley Initiative for Transparency in the Social Sciences. SK is funded by the National Health and Medical Research Council (APP1127932).

Author contributions

AC, HL and JM conceived the idea for the project and prepared the initial list of reporting items. All authors contributed to the project design and protocol development. AC and HL collected and analysed data with CLINVIVO. AC wrote the first draft of the manuscript. All authors provided substantive feedback on the manuscript and have read and approved the final version.

Acknowledgments

We thank Dr Rob Froud and the CLINVIVO team for providing their expertise and guidance in conducting and analysing this Delphi study. We are also grateful to the Delphi participants for their substantive engagement and feedback.

REFERENCES

- 1 Moore GF, Audrey S, Barker M, *et al.* Process evaluation of complex interventions: Medical Research Council guidance. *Br Med J* 2015;**350**:h1258. doi:10.1136/bmj.h1258
- 2 Lee H, Herbert RD, Lamb SE, *et al.* Investigating causal mechanisms in randomised controlled trials. *Trials* 2019;**20**:524. doi:10.1186/s13063-019-3593-z
- 3 Lange T, Hansen KW, Sørensen R, *et al.* Applied mediation analyses: a review and tutorial. *Epidemiol Health* 2017;**39**:e2017035. doi:10.4178/epih.e2017035
- 4 Lee H, Herbert RD, McAuley JH. Mediation Analysis. *JAMA*. 2019;**321**:697–8. doi:10.1001/jama.2018.21973
- 5 VanderWeele T. *Explanation in Causal Inference*. Oxford University Press 2015.
- 6 Keele L, Tingley D, Yamamoto T. Identifying mechanisms behind policy interventions via causal mediation analysis. *J Policy Anal Manag* 2015;**34**:937–63. doi:10.1002/pam.21853
- 7 Lee H, Lamb SE. Advancing Physical Therapist Interventions by Investigating Causal Mechanisms. *Phys Ther* 2017;**97**:1119–21. doi:10.1093/ptj/pzx095
- 8 Efficacy and Mechanism Evaluation | NIHR. <https://www.nihr.ac.uk/explore-nihr/funding-programmes/efficacy-and-mechanism-evaluation.htm> (accessed 28 Jan 2020).
- 9 Nguyen TQ, Schmid I, Stuart EA. Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. Published Online First: 2019.<http://arxiv.org/abs/1904.08515>
- 10 Liu S, Ulbricht CM, Chrysanthopoulou SA, *et al.* Implementation and reporting of causal mediation analysis in 2015: a systematic review in epidemiological studies. *BMC Res Notes* 2016;**9**. doi:10.1186/s13104-016-2163-7
- 11 Gelfand LA, Mensinger JL, Tenhave T. Mediation Analysis: A Retrospective Snapshot of Practice and More Recent Directions NIH Public Access. *J Gen Psychol* 2009;**136**:153–76.
- 12 Mansell G, Kamper SJ, Kent P. Why and how back pain interventions work: What can we do to find out? *Best Pract Res Clin Rheumatol* 2013;**27**:685–97. doi:10.1016/j.berh.2013.10.001
- 13 Hertzog M. Trends in Mediation Analysis in Nursing Research: Improving Current Practice. *West J Nurs Res* 2018;**40**:907–30. doi:10.1177/0193945916687782
- 14 Wood RE, Goodman JS, Beckmann N, *et al.* Mediation testing in management research: A review and proposals. *Organ Res Methods* 2008;**11**:270–95. doi:10.1177/1094428106297811
- 15 Lapointe-Shaw L, Bouck Z, Howell NA, *et al.* Mediation analysis with a time-to-event outcome: a review of use and reporting in healthcare research. *BMC Med Res Methodol* 2018;**18**:118. doi:10.1186/s12874-018-0578-7
- 16 Cashin AG, Lee H, Lamb SE, *et al.* An overview of systematic reviews found suboptimal reporting and methodological limitations of mediation studies investigating causal mechanisms. *J Clin Epidemiol* Published Online First: 20 March 2019. doi:10.1016/J.JCLINEPI.2019.03.005
- 17 Vo T, Superchi C, Boutron I, *et al.* The conduct and reporting of mediation analysis in recently published randomized controlled trials: results from a methodological systematic review. *J Clin Epidemiol* 2020;**117**:78–88. doi:10.1016/j.jclinepi.2019.10.001
- 18 Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010;**8**:18. doi:10.1186/1741-7015-8-18
- 19 Elm E von, Altman DG, Egger M, *et al.* Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;**335**:806–8. doi:10.1136/bmj.39335.541782.AD
- 20 Cashin AG, McAuley JH, Lamb SE, *et al.* Development of A Guideline for Reporting Mediation Analyses (AGReMA). *BMC Med Res Methodol* Published Online First: 2020. doi:10.1186/s12874-020-0915-5
- 21 Moher D, Schulz KF, Simera I, *et al.* Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;**7**:e1000217. doi:10.1371/journal.pmed.1000217
- 22 Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;**311**:376–80. doi:10.1136/BMJ.311.7001.376
- 23 Fitch K, Bernstein SJ, Aguilar MD, *et al.* The RAND/UCLA Appropriateness Method User's Manual Library of Congress Cataloging-in-Publication Data. Published Online First: 2001.

- doi:10.1180/minmag.2012.076.8.57
- 24 Froud R, Amundsen P, Bartys S, *et al.* Opportunities and challenges around adapting supported employment interventions for people with chronic low back pain: modified nominal group technique. *Disabil Rehabil* 2020.
 - 25 Froud R, Eldridge S, Kovacs F, *et al.* Reporting outcomes of back pain trials: A modified Delphi study. *Eur J Pain* 2011;**15**:1068–74. doi:10.1016/j.ejpain.2011.04.015
 - 26 Slade SC, Dionne CE, Underwood M, *et al.* Consensus on Exercise Reporting Template (CERT): Modified Delphi Study. *Phys Ther* 2016;**96**:1514–24. doi:10.2522/ptj.20150668
 - 27 Fink A, Kosecoff J, Chassin M, *et al.* Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;**74**:979–83. doi:10.2105/ajph.74.9.979
 - 28 Boulkedid R, Abdoul H, Loustau M, *et al.* Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLoS One* 2011;**6**:e20476. doi:10.1371/journal.pone.0020476
 - 29 Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Med* 2011;**8**:e1000393. doi:10.1371/journal.pmed.1000393
 - 30 Murphy MK, Sanderson C, Black NA, *et al.* Consensus development methods, and their use in clinical guideline development. 1998. www.hta.ac.uk/htacd.htm (accessed 27 Nov 2019).
 - 31 Staniszewska S, Brett J, Simera I, *et al.* GRIPP2 reporting checklists: Tools to improve reporting of patient and public involvement in research. *BMJ* 2017;**358**. doi:10.1136/bmj.j3453
 - 32 Simera I, Moher D, Hirst A, *et al.* Transparent and accurate reporting increases reliability, utility, and impact of your research: Reporting guidelines and the EQUATOR Network. *BMC Med* 2010;**8**:24. doi:10.1186/1741-7015-8-24
 - 33 Diamond IR, Grant RC, Feldman BM, *et al.* Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014;**67**:401–9. doi:10.1016/j.jclinepi.2013.12.002
 - 34 Pinnock H, Epiphaniou E, Sheikh A, *et al.* Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. *Implement Sci* 2015;**10**. doi:10.1186/s13012-015-0235-z

FIGURE LEGENDS

Figure 1. A simple mediation model. The total effect of the intervention/exposure on the outcome, represented by the broken line, is partitioned into an ‘indirect effect’ that works through the mediator of interest, and a ‘direct effect’ that works through all other unspecified mediators. The indirect effect is represented by the solid line from the intervention/exposure to the outcome through the mediator. The direct effect is represented by the solid line from the intervention/exposure to the outcome.[4]

Figure 2. Flow of reporting items through the Delphi study