BITSS SSMART Grant - Draft Report

Developing a Guideline for Reporting Mediation Analyses (AGReMA) in randomized trials and observational studies

Table of Contents

Project Background

Studies investigating causal mechanisms are becoming increasingly common. Despite growing numbers of publications and trialist embedding mechanism evaluations into randomised controlled trials and observational studies, the reporting accuracy and consistency of mechanism studies is suboptimal. The heterogeneity in the reporting of mechanism evaluations stifles systematic reviews, complicates meta-analyses, and limits transparency and replication. The aim of this initiative is to develop a reporting guideline for mechanism evaluations (mediation analyses). We also plan to develop an accompanying explanation and elaboration (E&E) paper. This project is the first initiative to address the issue of poor reporting quality of mechanism evaluations and has been registered on the "Enhancing the QUAlity and Transparency of health Research" (EQUATOR) network.

This initiative is being conducted in accordance with the Guidance for Developers of Health Research Reporting Guidelines. This draft report provides an update on the progress of the project; summarising achievements to date and future work. Currently, multiple phases of the overall program are running in parallel, with each phase informing the subsequent.

Phase 1 – Umbrella review of systematic reviews of mechanism evaluations

To improve research standards and publishing norms for mechanism evaluations, an examination of the current field and quality of reporting is needed. Therefore, we conducted an umbrella review to identify, assess and summarise all published peer-reviewed systematic reviews of mechanism evaluation studies across healthcare research over the previous 10 years, focusing on methodological and reporting elements specific to the studies reviewed.

Results (Overview)

Our electronic search yielded 2455 citations. Following stage 1 and 2 screening, 58 full text articles were included for review *(Figure 1)*.

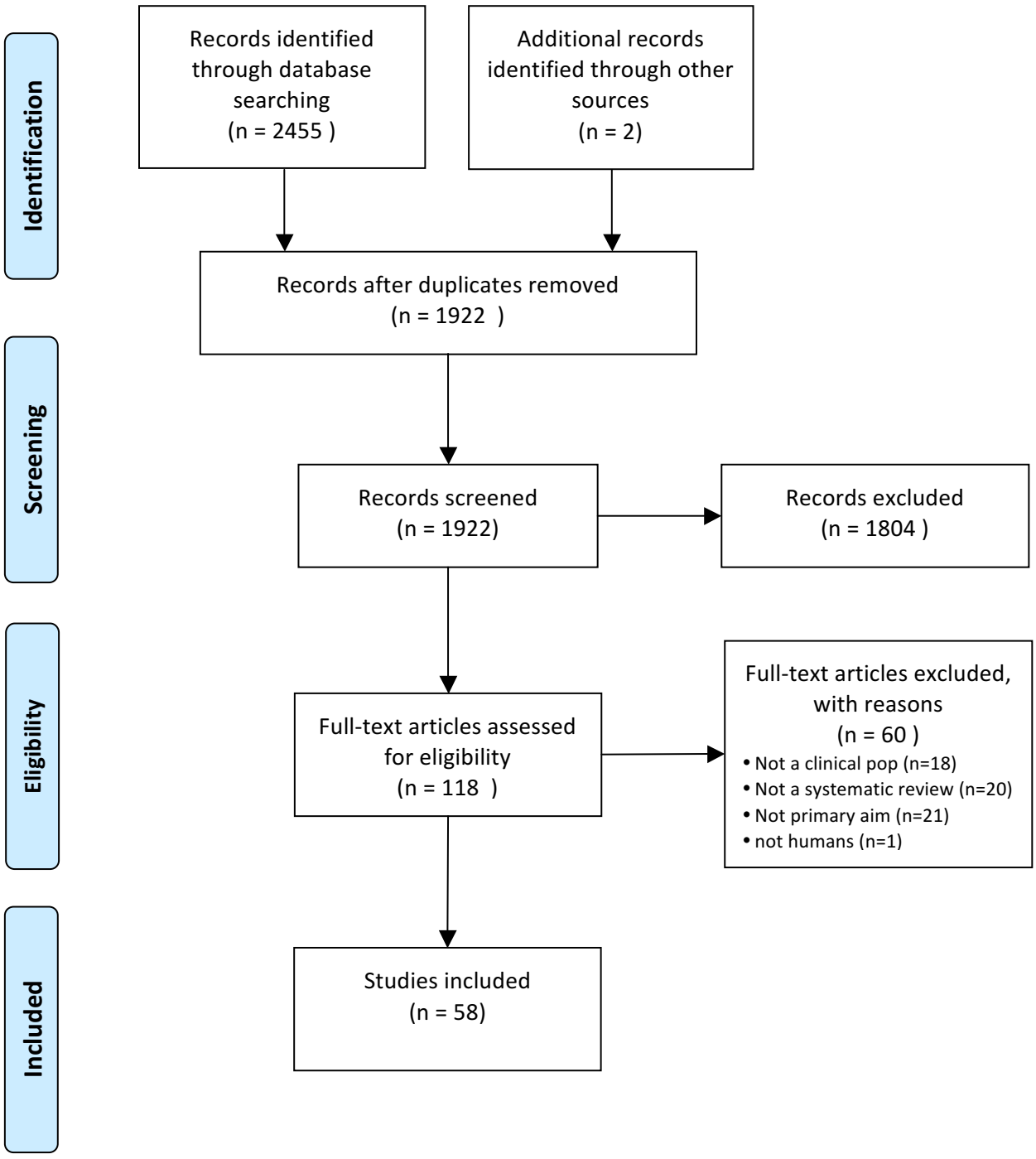


Figure 1. Study flow chart (adapted from PRISMA).

We found evidence for the need for improved *methodology* in mechanism evaluation studies and the need for more *transparent reporting*. These results support the need for the development of a reporting guideline for mechanism evaluation studies.

The following are some quotes from published reviews of mechanism evaluation studies:

> *"Reporting of causal mediation analysis is varied and suboptimal. Given that the application of causal mediation analysis will likely continue to increase, developing standards of reporting of causal mediation analysis in epidemiological research would be prudent." - Liu SH, Ulbricht CM, Chrysanthopoulou SA & Lapane KL. (2016). Implementation and reporting of causal mediation analysis in 2015: a systematic review in epidemiological studies. BMC Research Notes, 9, 354. doi:10.1186/s13104-016-2163-7*
>
> *"There was considerable variability in the reported statistical parameters (eg, R2, R2 change, partial r, standardized regression coefficients b) and the applied statistical methods...only 8 studies reported size and significance level of the indirect effect" - Schmidt, S. J., Mueller, D. R., & Roder, V. (2011). Social cognition as a mediator variable between neurocognition and functional outcome in schizophrenia: empirical review and new results by structural equation modeling. Schizophrenia bulletin, 37(suppl 2), S41-S54.*
>
> *"The evidence was insufficient to clearly determine the influence of predictors, moderators and mediators of intervention success due to the lack of consistent reporting across studies and the meta-regression being limited to the examination of age and gender (as the most commonly reported characteristics)." Miles, C. L., Pincus, T., Carnes, D., Homer, K. E., Taylor, S. J., Bremner, S. A., ... & Underwood, M. (2011). Can we identify how programmes aimed at promoting self-management in musculoskeletal pain work and who benefits? A systematic review of sub-group analysis within RCTs. European journal of pain, 15(8), 775-e1.*

Phase 2 - Systematic review of original mechanism evaluations in randomised controlled trials

Following the completion of phase 1, we have established a rationale for the need to develop a minimal but essential reporting guideline to improve the reporting standards of mechanism evaluation studies. To determine which items will be included in the Delphi rounds, we must first assess what is currently being reported or not reported. We will systematically review the current literature to identify randomised controlled trials embedding mechanism evaluations published in the last 5 years. The results from this study will inform the subsequent Delphi study where a

group of experts will reach consensus on the proposed items as well as be given the opportunity to provide recommendations regarding any item additions.

The protocol for the systematic review of original mechanism evaluation studies in randomised controlled trials has been developed and is in the process of submission for publication. The protocol is outlined below. Once the protocol is accepted, we will commence the literature search with results forecast to be available late 2017 to inform the commencement of the Delphi study.

Phase 3 - A modified Delphi

The results of phase 2 will inform the first round of the modified Delphi study. The Delphi technique is a structured method for consensus building among a panel of experts. It is characterised by a series of questionnaires or 'rounds' completed anonymously by participants followed by multiple iterations with controlled feedback. The feedback process employed allows and encourages participants to reassess their initial judgements and adjust their responses accordingly to group feedback in subsequent rounds. The Delphi technique is advantageous in that large numbers of participants across diverse locations can anonymously contribute avoiding instances where group discussions are dominated by the views of a few. We will develop a set of questionnaire items identified from the previous systematic review and perform a modified three-round Delphi process.

Our primary research question is: "What items should be included in a minimal but essential reporting guideline for mechanism evaluations?". The protocol for the modified Delphi study is seen in the appendix.

We have constructed a preliminary list of experts to participate in the modified Delphi. This list was informed by authors of systematic reviews of mechanism evaluation studies reviewed in Phase 1. 65 potential "expert" participants have been identified from a mixture of professional backgrounds (i.e., journal editor, methodologist, academic, clinician). Table 1 details the geographic distribution of identified experts.

Table 1. Geographic distribution of experts

Country	No	%
Australia	7	10.76923077
Belguim	1	1.538461538
Denmark	4	6.153846154
Germany	4	6.153846154
Netherlands	7	10.76923077
Hong Kong	2	3.076923077
Portugal	2	3.076923077
Switzerland	6	9.230769231
United Kingdom	15	23.07692308
United States of America	17	26.15384615

Ethics for the modified Delphi study has been obtained by the Human Research Ethics Advisory (HREA) Panel D: Biomedical, The University of New South Wales, NSW, Australia.

Project title: Developing a Guideline for Reporting Mediation Analysese (AGReMA) – a modified Delphi study.
HC No: HC16599
Approval Period: 31-Aug-2016 – 30-Aug-2021

A proposed list of preliminary items to be included in the Delphi has been constructed informed from the results of the Phase 1 review – Table 2. This will be corroborated with findings from Phase 2.

Table 2. Preliminary items to be included in the Delphi

Section/Topic	Checklist item
Title and abstract	
	Identification as a mechanism evaluation / mediation analysis in the title
Introduction	
Background and objectives	Scientific background, including cited theoretical framework for proposed mediators, and explanation of rationale for mechanism evaluation
	Specific objectives or hypotheses for mechanism evaluation
Methods	
Trial design	Description of trial design including how the design features enable mediation
	Description of measures to control for possible confounding factors given specific mediation question and trial design
	Explanation of why the intervention influences proposed mediating variables

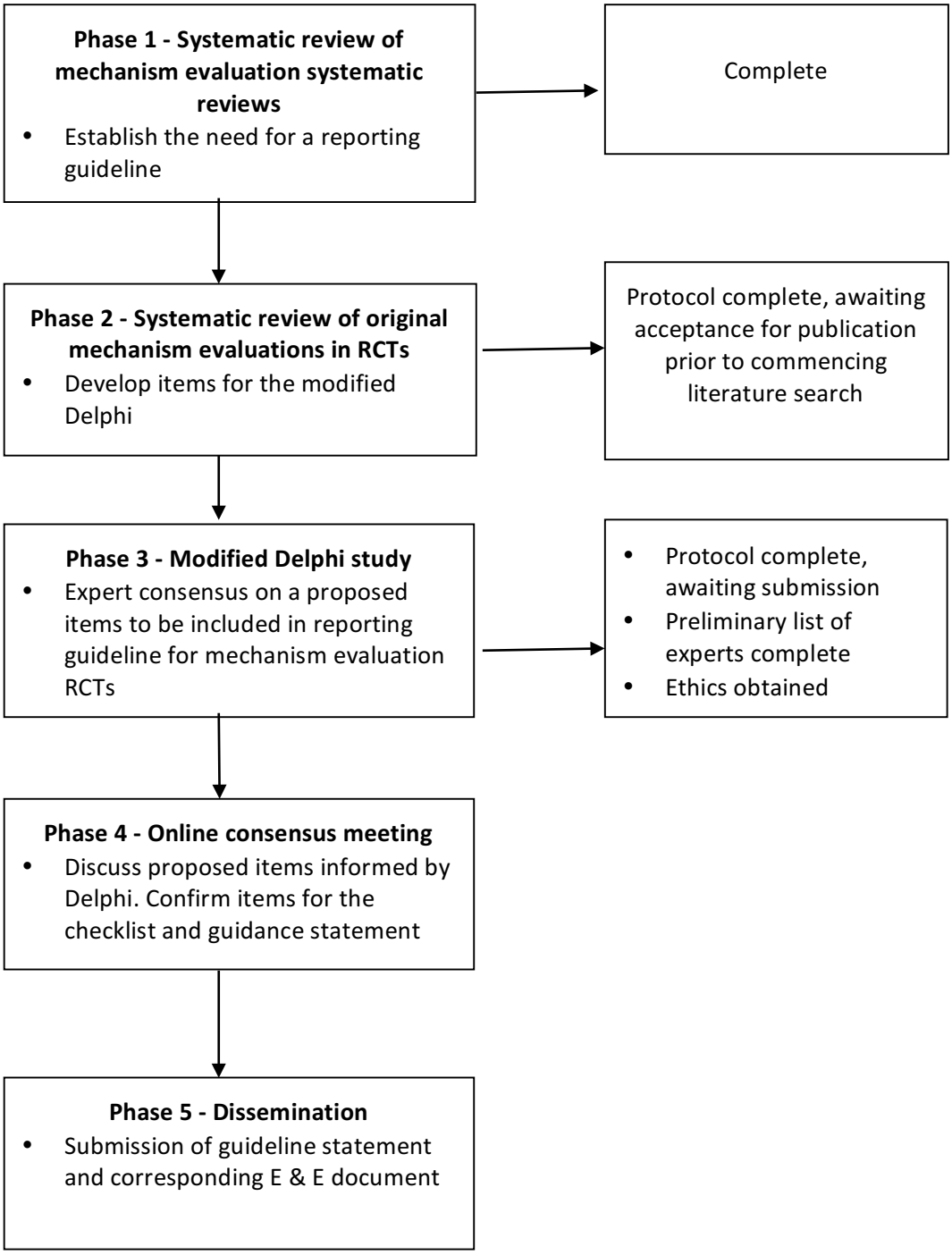
Outcomes	Completely defined pre-specified primary outcome measures, including how and when they were assessed and their psychometric characteristics (within acceptable ranges i.e., test-retest or Cronbach's alpha>.60)
	Completely defined pre-specified proposed mediator variable measures, including how and when they were assessed and their psychometric characteristics (within acceptable ranges i.e., test-retest or Cronbach's alpha>.60)
	If identified, completely defined pre-specified suppression variables, including how and when they were assessed and their psychometric characteristics (within acceptable ranges i.e., test-retest or Cronbach's alpha>.60)
	Clearly stated temporal sequence between outcome and mediator variables
	Any changes to trial outcomes or mediator variables after the trial commenced, with reasons
Sample size	How sample size was determined including power calculation to detect indirect effects
Statistical methods	Statistical methods used to assess mediation
	Statistical methods for satisfying the “sequential ignorability” assumption i.e., sensitivity analysis
	Statistical methods to explicitly test the mediated effect i.e., Sobel test, bootstrapping etc.
	Software and packages used for statistical tests
Results	
Treatment compliance	Outline % treatment compliance and assumptions met to estimate mediation effects following significant treatment noncompliance
Outcomes and estimation	For each primary outcome, results for each group, and the estimated total effect size and its precision (such as 95% confidence interval)
	Results for the direct effect reported in the raw metric of analysis with its precision (such as 95% confidence interval)
	Results for the indirect effect (Paths a and b) reported in the raw metric of analysis with its precision (such as 95% confidence interval)
	For any identified suppression effects, report the raw metric of analysis and its precision (such as 95% confidence interval)
	Results from sensitivity analysis – for example - how the violation of the sequential ignorability assumption invalidates the indirect effect
	Table for regression model association (95% CI and p-Value) between treatment potential mediators - path a
	Table for indirect, direct and total effects of the mediation model with primary outcome at test time points

Summary

The development of this reporting guideline should improve standardisation, accuracy, transparency, and completeness in reporting mechanism evaluations. This will facilitate more comprehensive systematic reviews of mechanisms, and guide researchers, editors, and policy makers in evaluating evidence for causal mechanisms. The first 3 phases of this initiative have commenced and are running in

parallel. The completion of the project will involve the dissemination and implementation of the guideline, consisting of simultaneous publications and requests to journal editors to endorse the guideline. The results will also be presented by the project executive at international conferences, professional bodies, and organisations across multiple disciplines phase. Figure 4 summarises the 5 phases of this initiative.

Figure 4 - Flow Diagram depicting project phases



Appendix

Phase 2 - Exploring the quality of reporting of mechanism evaluations in randomised controlled trials: protocol for a systematic review

INTRODUCTION

Randomised controlled trials (RCTs), when appropriately designed, conducted, and reported, are considered the gold standard for evaluating the efficacy and effectiveness of healthcare interventions.[1] However, many of these empirical studies focus on *whether* and to what extent, one variable effects and another.[2] Of late, there has been a surge of interest to go beyond this "black-box" approach to causality and unpack the effects of healthcare interventions, to ask, *how* or *why* certain variables causally affect the outcome.[2,3] A standard analytical method of evaluating causal mechanisms is via mediation analysis.[4] Mediation analysis tests the inclusion of a third intermediate variable or mediator (M), on the causal pathway between an exposure (X) and an outcome (Y)(Figure 2).[2] For example, an education intervention (X) aimed to reduce health-care utilisation (Y), could be mediated by changes in the level of patient self-efficacy (M). The identification of a causal mechanism and the integration of this information into implementation strategies and existing health services, presents a refined method of redesigning healthcare and improving patient management.

Studies investigating causal mechanisms from RCTs are becoming increasingly common.[5,6] Despite growing numbers of publications and trialist embedding mechanism evaluations into RCTs, there remains some confusion regarding appropriate methodology and adequate reporting.[3,5] Across healthcare research, systematic reviews assessing original published mechanism studies note inconsistency in methodological design, statistical approaches, and reporting standards.[7–11] The heterogeneity in the reporting of mechanism evaluations stifles systematic reviews, complicates meta-analyses, and limits transparency and replication.

Reporting standards have been encouraged for randomised controlled trials since the development of the original CONSORT (Consolidated Standards of Reporting Trials) statement in 1996.[12] Currently, there are no relevant reporting guidelines for mechanism evaluation studies registered on the "Enhancing the QUAlity and Transparency of health Research" (EQUATOR) network. As part of an initiative to develop a reporting guideline for mechanism evaluations in RCTs, we will first systematically review the current literature selecting a sample of randomised controlled trials embedding mechanism evaluations between 2014 and 2016, focusing on the reporting elements specific to those trials.

The aim of this systematic review is to summarise the current state of reporting for mechanism evaluations in randomised controlled trials and secondly, to highlight key reporting items that will inform the development a minimal but essential reporting guideline for mechanism evaluations in randomised controlled trials.

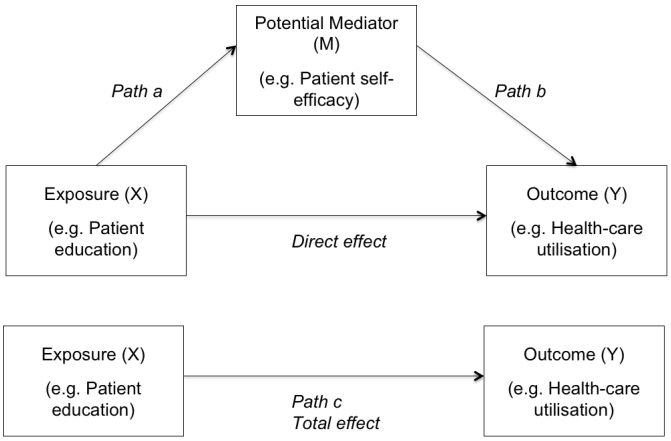


Figure 2. Single mediator model. The indirect effect is quantified by the product of paths a and b. The total effect (path c) is the sum of the indirect and direct effects.

METHODS

Search Strategy and Study Selection

We will perform a computerised search of the MEDLINE databases and the Cochrane Central Register of Controlled Trials using the search terms *mechanism evaluation OR mediat* OR mediation analysis OR mediation analyses OR causal mediation analysis OR causal mediation analyses OR mechanism evaluation AND random**. We will identify all reports of randomised controlled trials on human healthcare interventions published in the English language, from January 1st, 2014 to December 31st, 2016. Two reviewers (AC and HL) will screen the titles and abstracts to identify potentially relevant studies, and the final selection will be made from reading the full text. We will resolve discrepancies through discussion, and a third author (JM) will adjudicate unresolved disagreements where necessary. Articles will be included in the study if identified as a randomised controlled trial assessing causal mechanisms through a formal statistical mediation analysis in a healthcare intervention. Letters and articles for which full text is unavailable or reports describing only the design or protocol of the trial will be excluded. Articles will be screened for duplicate publication (ie, the same trial described in several articles), and only the article presenting the main results will be included. We will record the reasons for excluding trials. This systematic review protocol is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols 2015 (PRISMA-P 2015) statement. [13] In accordance with the guidelines, our systematic review protocol will be registered with the International Prospective Register of Systematic Reviews (PROSPERO).

Evaluation of Methodological Quality

Two independent reviewers (AC and HL) will develop and test a data extraction form with a distinct set of 10 articles during a training session. Following this, the reviewers will meet and discuss the interpretation of different items. To assess interobserver reliability, we will conduct a calibration exercise of a subsample of 30 randomly selected articles, and the two reviewers will independently extract the information. For the remaining articles, a single reviewer (AC) will extract information. Reviewers will not be blind to the journal name or authors.

The data extracted will include:

1. Report of the study characteristics including year of publication, number of patients, type of intervention, number of groups, level of blinding, randomisation process, whether the choice and rationale for mediation analysis was clearly detailed.
2. Report of the study design including cited theoretical framework for proposed mediating variables, previous reported pilot studies to test the intervention effect on

mediators and whether the study measures/procedures were designed to influence the mediating variables.
3. Report of all elements of the sample size calculation including whether the study was adequately powered to detect meditation and the presence of type 1 and type 2 error rates, common standard deviation and difference between groups.
4. Report of psychometric characteristics of mediating and outcome variables and if they were within accepted ranges (Cronbach's alpha and test-retest reliability >.60).
5. Established temporal precedence through the reporting of variable measurement time frames to ascertain whether changes in mediating variables precede changes in outcome variables.
6. Method of statistical testing used. We will note whether the mediation analysis was priori planned and any observed variation. We will also determine whether the results were presented with CIs or P values.
7. Report of mediation outcome. Did the change in the potential mediator correlate with change in outcome.
8. Report of possible confounding factors and underlying assumptions. We will note methods use to adjust for such assumptions and whether sensitivity analysis was performed.

Statistical Analysis

Categorical variables will be described with frequencies and percentages. The degree of agreement between the 2 reviewers will be determined using the k coefficient. Analyses will be performed in R (The R Foundation for Statistical Computing).

Conclusion

With increasing publication of mechanism evaluations across multiple disciplines, it is likely that more systematic reviews of mechanism evaluations will be conducted. The development and implementation of a minimal but essential reporting guideline for mechanism evaluations in randomised controlled trials has the potential to improve research standards, scientific reporting and provide greater transparency on a large scale.

References

1. Schulz KF, Altman DG, Moher D. Open Access CORRESPONDENCE CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med Obstet Gynecol PLoS Med Ann Intern Med Open Med J Clin Epidemiol BMC Med Trials*. 2010;8. doi:10.1186/1741-7015-8-18.
2. Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *Am Polit Sci Rev*. 2011;105(4):765-789. doi:10.1017/S0003055411000414.
3. Mackinnon DP, Luecken LJ. How and for Whom? Mediation and Moderation in Health Psychology. doi:10.1037/0278-6133.27.2(Suppl.).S99.
4. Mansell G, Hill JC, Kamper SJ, Kent P, Main C, van der Windt D a. How can we design low back pain intervention studies to better explain the effects of treatment? *Spine (Phila Pa 1976)*. 2014;39(5):E305-310. doi:10.1097/BRS.0000000000000144.
5. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*. 2015;350:h1258-h1258. doi:10.1136/bmj.h1258.
6. Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. *J R Stat Soc*. 2008;171(2):481-502.
7. Lee H, Hübscher M, Moseley GL, et al. How does pain lead to disability? A systematic review and meta-analysis of mediation studies in people with back and neck pain. *Pain*. 2015;156(6):988-997. doi:10.1136/bjsports-2013-092930.
8. Lubans DR, Foster C, Biddle SJH. A review of mediators of behavior in interventions

to promote physical activity among children and adolescents. *Prev Med (Baltim)*. 2008;47(5):463-470. doi:10.1016/j.ypmed.2008.07.011.
9. 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;27:685-697. doi:10.1016/j.berh.2013.10.001.
10. Milesl CL, Pincusl T, Carnesl D, et al. Review: Can we identify how programmes aimed at promoting self-management in musculoskeletal pain work and who benefits? A systematic review of sub-group analysis within RCTs. *Eur J Pain*. 2011;15(8):775.e1-775.e11. doi:10.1016/j.ejpain.2011.01.016.
11. Schmidt SJ, Mueller DR, Roder V. Social cognition as a mediator variable between neurocognition and functional outcome in schizophrenia: Empirical review and new results by structural equation modeling. *Schizophr Bull*. 2011;37(SUPPL. 2). doi:10.1093/schbul/sbr079.
12. Begg C, Cho M, Eastwood S, et al. Improving the Quality of Reporting of Randomized Controlled Trials. *JAMA*. 1996;276(8):637. doi:10.1001/jama.1996.03540080059030.
13. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. doi:10.1186/2046-4053-4-1.

Phase 3 - Developing a reporting guideline for mechanism evaluations in randomised controlled trials: protocol for a modified Delphi study

INTRODUCTION

The primary research objective for most empirical social and medical scientists is the identification and estimation of causal effects. That is, *whether* a particular exposure causally affects an outcome. More recently, scientists have gone beyond this question to ask, *how* certain variables causally affect the outcome. This line of research goes beyond the 'black-box approach' to identifying causal effects and extends the research capacity to ask pertinent questions about underlying causal mechanisms. A standard analytical method of evaluating causal mechanisms is via mediation analysis. This methodology is now becoming a common part of many randomised trials.[1,2]

Mediation analysis is used to test and quantify theories regarding causal links between an exposure and an outcome.[3] Alongside examining the direct effect between the exposure and outcome, mediation analysis examines the extent to which an intermediate variable (mediator) explains the effect[4,5] (Figure 3). Mediation analysis can be applied to data from various types of study design: cross-sectional surveys, clinical registries, longitudinal cohorts and randomised and non-randomised clinic trials.[3] Although different assumptions are presented with different study designs impacting the overall explanatory power, the flexibility and utility of this approach makes mediation analysis a useful supplement to other, more commonly used, methods of analysis.[3]

Despite widespread use, the reporting accuracy and consistency of mechanism evaluations is not consistent. Published mechanism studies omit critical components of methodological design, statistical approaches, and report a variety of effect estimates. Poorly reported mechanism studies stifles replication, limits risk of bias assessments, complicates meta-analyses, wastes resources, and is ultimately unethical. There is currently no published initiative to develop a reporting guideline for mechanism studies, and there are no relevant guidelines registered on the "Enhancing the QUAlity and Transparency of health Research"

(EQUATOR) network.

There is a need to develop a new reporting guideline for mechanism evaluations in order to improve the quality of reporting standards in randomised controlled trials. The implementation of the proposed tool will instigate clearer documentation and reporting practices, thereby allowing for more complete meta-analyses with reduced likelihood of bias. Involving international experts in the field of mechanism analysis via a modified-Delphi technique, will assist the selection and refinement of items for a new reporting guideline.

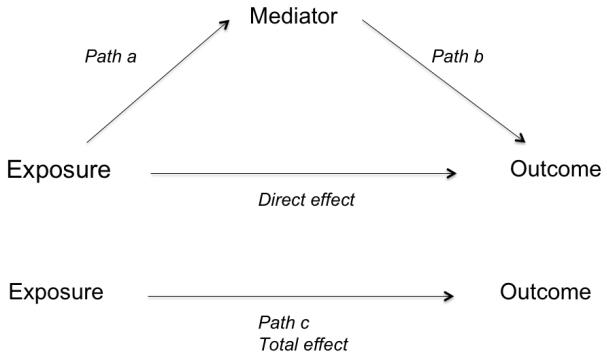


Figure 3. Single mediator model. The indirect effect is quantified by the product of paths a and b. The total effect (path c) is the sum of the indirect and direct effects.

AIMS

We will develop a reporting guideline for mechanism evaluations in order to improve the quality of reporting standards in randomised controlled trials using international expert consensus in a modified-Delphi technique. We will follow the recommended 18-step check list for best practice developing a reporting guideline by Moher *et al*[6] for the EQUATOR Network.
Our primary research question is: "What items should be included in a minimal but essential reporting guideline for mechanism evaluations?".

METHODS AND DESIGN

The Delphi technique is a structured method for consensus building among a panel of experts. It is characterised by a series of questionnaires or 'rounds' completed anonymously by participants followed by multiple iterations with controlled feedback. The feedback process employed allows and encourages participants to reassess their initial judgements and adjust their responses accordingly to group feedback in subsequent rounds.[7] The Delphi technique is advantageous in that large numbers of participants across diverse locations can anonymously contribute avoiding instances where group discussions are dominated by the views of a few.[8,9]
We will develop a set of questionnaire items identified from recent systematic reviews and perform a modified three-round Delphi process. Consensus will be defined as >70% agreement on items considered high importance to be included in the reporting guidelines. We will align our study with the proposed key methodological quality indicators for Delphi studies: reproducible criteria for participant selection, transparent study aim and consensus definition, planned number of rounds, clearly defined stopping criteria and criteria for item removal/inclusion.[10]

Sampling

Participants will involve "experts" in mechanism evaluations - defined as researchers who have contributed to the literature by: conducting original mediation analyses; conducing systematic reviews of original mechanism evaluation studies; and those who have contributed to the methodological and statistical advance of mediation analysis. Potential participants will be identified through an iterative process. First, by compiling a list of potential participants informed by editorial boards, publications, conference proceedings, and organisational memberships. Subsequently, a "snowball sampling" approach will be used to allow identified participants to nominate further potential participants. To be eligible for the study participants have to: hold a current appointment with an academic institute, have published at least 1 original research paper using mediation analysis or a methodological/statistical research paper on mechanism evaluations, be an author of textbooks on mediation analysis, be a journal editors serving on the board of a journal that publishes mechanism evaluations, have a sufficient understanding of written and spoken English language, be aged over 18 years. Exclusion criteria includes: those who have only conducted "Process evaluations" without a statistical mediation analysis.

Data collection

We will use SurveyMonkey (http://surveymonkey.com) to recruit potential participants and to conduct the online Delphi process. SurveryMonkey is an ideal online platform for conducting and analysing input from geographically dispersed participants. The recruitment survey will include an explanatory statement about the project and collect demographic information on: age, gender, discipline, current practice and experience in mechanism evaluations.

We will use a modified three-round Delphi process to identify the most important items for inclusion in a minimal but essential reporting guideline for mechanism evaluations in randomised controlled trials. Each survey round will be online for up to 4 weeks and reminder emails will be sent approximately every 7 days after the initial invitation. Rounds will be progressed when response rates >60%. In round one, participants will be asked to rate items on their importance for inclusion using a 9-point Likert scale (anchored by 1 - "not at all important for inclusion" and 9 - "essential for inclusion"). Participants will also be asked to provide their rationale for their ratings in free-text boxes under each item.

In round two, all participants will be presented with visual plots and results tables from round one, including: median, inter-quartile range, frequencies for each item, and the participant's rating compared to group ratings. We will also show group's agreement/disagreement scores determined by inter-percentile range adjusted for symmetry (IPRAS) analysis technique from the RAND/UCLA Appropriateness method.[13] If an item has disagreement, it is considered to have uncertain importance – this will be explored in round two. If an item has agreement, the tertile in which the median rating for importance falls will be presented: median between 1 and 3 = low importance, 4 and 6 = moderate importance, and 7 and 9 = high importance. Participant comments from round one will also be provided next to each item. After presentation of these results, participants will be asked to rate all ambiguous items or proposals driven by comments of the first round and concerning exclusion, aggregation or retention of items, together with any new potential items identified from the first round.

In round three, we will include all items that have reached consensus for inclusion in rounds 1 and 2 in their original format, items that reached consensus for inclusion in round 2 but required further clarification, and any remaining items for which no consensus had been reached. This will allow participants to revise their initial views and re-identify items they think are important for the reporting guideline.

Ethics

Ethics has been obtained by the Human Research Ethics Advisory (HREA) Panel D: Biomedical, The University of New South Wales, NSW, Australia. Potential participants will be informed that by responding to the recruitment survey, they will be deemed to have given informed consent to participate in the study and have their de-identified responses included in any analysis. Participant anonymity will be maintained through the Delphi process through the use of usernames in the surveymonkey online platform (e.g., Participant 1, Participant 2, etc.). All data will be stored on a computer which is password encrypted, in a locked office, in accordance with standard guidelines. Only the researches will have access to the data which will be destroyed after 5 years in accordance with standard guidelines.

Analysis

A descriptive analysis of item ratings will summarise the distribution of group responses from each round, describe changes in-group responses between rounds, and determine items that have agreement/disagreement. Items that reach consensus for high importance (using the IPRAS method) in round three will be included in the reporting guideline for mechanism evaluations of randomised controlled trials. This analysis procedure follows similar approaches to that used in the development of a reporting guideline for trial protocols of social science interventions. The IPRAS method is described below:

Agreement/disagreement for each item will be calculated with the following equations:

1. Determine the lower limit of the 40% inter-percentile range (IPRL; 30th-percentile score)
2. Determine the upper limit of the 40% inter-percentile range (IPRU; 70th-percentile score)
3. Determine the central point of the IPR (IPRC): (IPRL + IPRU)/2
4. Determine the “Asymmetry Index” (AI) on the 9-point Likert scale: 5 - IPRC
5. Calculate the IPRAS for the item: 2.35 + (1.5*AI)
6. Calculate the IPR: IPRU - IPRL
7. Calculate the “Disagreement Index” (DI): IPR/IPRAS
8. Determine whether disagreement exists: DI > 1

DISCUSSION

The Delphi study will identify a priority list of items to inform the reporting guideline. This methodological tool –ie., reporting guideline- will strengthen the reliability and reporting accuracy of mechanism evaluations. With increasing publication of mechanism evaluations across multiple disciplines, it is likely that more systematic reviews of mechanism evaluations will be conducted. The development and implementation of this new methodological tool will change scientific reporting and provide transparency in investigating causal mechanisms in randomised controlled trials. This project has the potential to improve research standards and publishing norms on a large scale.

DISSEMINATION PLAN

The findings of this study will inform the development of the reporting guideline and the Explanation and Elaboration (E&E) document. We will disseminate and implement the guidelines through simultaneous publications in peer-reviewed journals and requests to journal editors to endorse the guideline. The project executive will present the guideline at international conferences, professional bodies, and organisations across multiple disciplines.

References

1. Moore, G. F. *et al.* Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* **350,** h1258–h1258 (2015).

2. Imai, K., King, G. & Stuart, E. A. Misunderstandings between experimentalists and observationalists about causal inference. *J. R. Stat. Soc.* **171,** 481–502 (2008).

3. Mansell, G., Kamper, S. J. & Kent, P. Why and how back pain interventions work: What can we do to find out? *Best Pract. Res. Clin. Rheumatol.* **27,** 685–697 (2013).

4. Mackinnon, D. P., Fairchild, A. J. & Fritz, M. S. Mediation Analysis. doi:10.1146/annurev.psych.58.110405.085542

5. Lee, H. *et al.* How does pain lead to disability? A systematic review and meta-analysis of mediation studies in people with back and neck pain. *Pain* **156,** 988–997 (2015).

6. Moher, D. *et al.* Guidance for Developers of Health Research Reporting Guidelines. *PLoS Med.* **7,** e1000217 (2010).

7. Hsu, C. & Sandford, B. The delphi technique: making sense of consensus. *Pract. Assessment, Res. Eval.* **12,** 1–8 (2007).

8. Boulkedid, R. *et al.* Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLoS One* **6,** e20476 (2011).

9. Sinha, I. P. *et al.* 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.* **8,** e1000393 (2011).

10. Diamond, I. R. *et al.* Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J. Clin. Epidemiol.* **67,** 401–409 (2014).

11. SurveyMonkey, L. L. C. SurveyMonkey®. (2012).

12. Khodyakov, D. *et al.* Conducting Online Expert panels: a feasibility and experimental replicability study. *BMC Med. Res. Methodol.* **11,** 174 (2011).

13. Fitch, K., Bernstein, S. J., Aguilar, M. D., Burnand, B. & LaCalle, J. R. *The RAND/UCLA appropriateness method user's manual*. (DTIC Document, 2001).