BMJ Open

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017972
Article Type:	Protocol
Date Submitted by the Author:	29-May-2017
Complete List of Authors:	Khadjesari, Zarnie; King's College London, Department of Health Service and Population Research Vitoratou, S; King's College London, Biostatistics and Health Informatics Department Sevdalis, Nick; King's College London, Department of Health Service and Population Research Hull, Louise; King's College London, Department of Health Service and Population Research
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Evidence based practice, Health services research
Keywords:	systematic review, implementation outcomes, implementation science, measurement properties, psychometric properties

SCHOLARONE™
Manuscripts

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

Zarnie Khadjesari*, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, 16 De Crespigny Park, Camberwell, London SE5 8AF, United Kingdom
zarnie.khadjesari@kcl.ac.uk

Silia Vitoratou, Psychometrics and Measurement Lab, Biostatistics and Health Informatics Department, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
silia.vitoratou@kcl.ac.uk

Nick Sevdalis, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
nick.sevdalis@kcl.ac.uk

Louise Hull, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
lousie.hull@kcl.ac.uk

*Corresponding author

ABSTRACT

Introduction

Over the past 10 years, research into methods that promote the uptake, implementation and sustainability of evidence-based interventions has gathered pace. Implementation outcomes, however, are defined in different ways and assessed by different measures, making it difficult to evaluate the effectiveness of implementation strategies and ultimately implementation success. The aim of this systematic review is to identify and appraise studies that assess the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings, to advance the use of precise and accurate measures.

Methods and analysis

The following databases will be searched for published literature from database inception to March 2017: MEDLINE, EMBASE, PsycINFO, CINAHL, HMIC and the Cochrane Library. Grey literature will be sought via OpenGrey, ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science. Reference lists of included studies and relevant reviews will be hand searched. Three search strings will be combined to identify eligible studies: 1) implementation literature, 2) implementation outcomes, 3) measurement properties. Screening of titles, abstracts and full papers will be assessed for eligibility by two reviewers independently and any discrepancies resolved via consensus with the wider research team. The methodological quality of the studies will be assessed using the COSMIN checklist. A set of bespoke criteria to determine the quality of the instruments will be used, and the relationship between instrument usability and quality will be explored.

Dissemination

Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of the most suitable implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Systematic review registration

International Prospective Register of Systematic Reviews (PROSPERO): CRD42017065348

Key words:

Systematic review; protocol; implementation outcomes; measurement properties; psychometric properties, implementation science

2

Strengths of this study

- We have designed a comprehensive search strategy for published and unpublished literature, and include a string of search terms for the type of measurement property.
- This will be the first systematic review of implementation outcomes that assesses the methodological quality of included studies.

Limitations of this study

- Due to the breadth of the setting (i.e. all physical healthcare settings), a validated search filter for measurement properties was not suitable as our approach needed greater precision for screening to be manageable.
- We selected a taxonomy of implementation outcomes to guide the selection of implementation outcomes in this review, however, there are several other models, theories and frameworks could have guided the identification of measures in this field.

3

BACKGROUND

Routinely delivered, evidence-based practice is a principal objective of healthcare systems across the world. However, the so called 'evidence-to-practice gap' means it can take many years before patients benefit from evidence-based interventions, if at all, and when implementation is attempted it is often fraught with barriers (1). Over the past 10 years, research into methods that promote the uptake of evidence-based practices (i.e. implementation research) has substantially increased (2). However, due to the emerging state of the field and the breadth of disciplines it covers, implementation outcomes are defined in different ways and assessed by a variety of different measures, making it difficult to evaluate and compare the effectiveness of different implementation strategies- "methods or techniques used to enhance the adoption, implementation, and sustainability of a clinical programme or practice" (3–5). Implementation outcomes reflect the impact of efforts to implement evidence-based treatments, practices and services and are distinct from service and client/patient outcomes, which are essential but not sufficient for understanding implementation success, or failure (6). As such, it has been argued that implementation outcomes should be defined and measured in all studies of implementation (7). It has been proposed that implementation outcomes serve three functions: 1) indicate *implementation success*, which is a prerequisite for the effectiveness of treatment and quality of care approaches; 2) constitute proximal indicators of implementation processes; 3) provide important intermediate outcomes for service and client/patient outcomes (7). Accurate and precise measurement of implementation outcomes is thus vital for developing the evidence-base on effective implementation strategies (8).

Previous reviews have focussed on measures of system level antecedents to implementation (9), organisational level culture and readiness to change (10–12), and individual level determinants of research utilisation (13), as well as predictors of innovation adoption (14). Chaudoir et al identified 61 instruments that predict implementation of evidence-based interventions at multiple levels, with the majority assessing organisation, provider, and innovation-level constructs, as opposed to structural or patient-level constructs (15). More recently, reviews have taken a broader approach and identified instruments that assess the 37 constructs contained in the Consolidated Framework of Implementation Research (CFIR)- a meta-theoretical framework that aims to understand and/or explain influences on implementation outcomes (16–18). Furthermore, a review has focussed on identifying quantitative measures of the eight implementation outcomes included in Proctor et al's working taxonomy (17). Lewis et al identified 104 instruments that measure these constructs in mental healthcare settings: the vast majority of the instruments measured acceptability (n=50), followed by adoption (n=19), feasibility (n=8), cost (n=8), sustainability

4

(n=8), appropriateness (n=7) and penetration (n=4). The review highlighted that implementation outcome instrumentation is underdeveloped with regards to the number of instruments available and the measurement quality of instruments.

This systematic review will use Proctor et al's working taxonomy of implementation outcomes to guide the identification of implementation outcome instruments used in physical healthcare settings. The working taxonomy of implementation outcomes is relevant across stakeholder levels, stages of implementation, and can applied to different implementation models, theories and frameworks (19). This review will complement and allow direct comparison of the review of Lewis et al, whose review used the taxonomy to identify instruments used in mental health settings (17), where instruments were largely found to be specific to a particular intervention, behaviour and/or setting, to provide a complete picture of all available measures and their properties.

Systematic reviews of measurement properties have been found to include non-comprehensive search strategies and lack methodological quality assessment, which are fundamental components of this research methodology, i.e. identifying all relevant literature in a field and providing information on the extent to which study results may be biased. A systematic review on the quality of systematic reviews of health-related outcome measurement instruments identified 102 reviews in a one-year period, and found only 59% had searched EMBASE (whilst searching MEDLINE and EMBASE databases is considered a minimal requirement), 54% did not include search terms for measurement properties, and only 41% assessed the methodological quality of the studies (20).

This systematic review will address the methodological limitations of earlier reviews, namely, it will use a comprehensive search strategy, and it will assess the methodological quality of the included studies using the COSMIN checklist (21), which in-turn will inform the assessment of the instruments quality. In using a similar methodological approach to the Lewis et al review, we can compare our findings with those from the mental health field in terms of the methodological quality of the studies (the COSMIN will be applied to an update of the mental health review), the psychometric quality of the instruments for each outcome, and the impact of usability on the psychometric quality of the instruments- where pragmatic/usable measures are vital for the implementation of the instruments themselves (22). The purpose of this review is to promote and advance the use of precise and accurate measures of implementation outcomes across all physical healthcare settings.

5

METHODS

This review protocol has followed the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P) 2015 checklist (23,24).

Aim

- To systematically identify and appraise studies that assess the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings.

Objectives

- To systematically identify studies that assess the measurement properties of quantitative implementation outcome instruments
- To critically appraise the methodological quality of the evidence on measurement properties of implementation outcome measures using the COSMIN checklist
- To apply a bespoke criteria to determine the psychometric quality of the instruments
- To explore the relationship between instrument usability and quality.

Stakeholder group

This protocol has been developed with the support of an international stakeholder group, whose role is to ensure the research conducted by the Centre for Implementation Science, King's College London (where the review team are based) is of direct relevance to stakeholders' needs. The group consists of healthcare professionals, managers and academics working in the field of implementation science including journal editors and grant panel members. We have also received feedback on the protocol from the Centre for Implementation Science and King's Improvement Science research teams.

Search strategy

Three sets of search terms will be combined to identify studies that assess the measurement properties of instruments that measure implementation outcomes. The search strings describe: 1) the population / field of interest (i.e. implementation literature), 2) the constructs being measured (e.g. adoption) and 3) the measurement properties of instruments (e.g. test-retest reliability) (25). The first string of terms will be used to identify the implementation literature (such as implement* OR knowledge transfer), incorporating terms used by Lewis et al (26), the UK Health Foundation's scoping review on the concept and practice of improvement science (27), and index terms (e.g. MeSH) applied to Lewis et al's published systematic review protocol and publication of findings. The second string of terms will consist of the implementation outcomes included in Proctor et al's taxonomy and their

6

synonyms (7,26). The third string of terms will relate to specific measurement properties of the instruments (such as internal consistency and content validity) (see Table 1).

We reviewed these search terms with our stakeholder groups to ensure they included all relevant synonyms. We will also conduct a supplementary search for the names of the instruments which are identified as eligible for inclusion in the review.

Published literature search

The following electronic databases will be searched using the search terms outlined above: MEDLINE, EMBASE, PsycINFO and HMIC (Health management Information Consortium) via the Ovid interface; CINAHL via the EBSCO Host interface; and the Cochrane library. Databases will be searched from inception to March 2017, there will be no language restrictions, a filter for studies in humans will be applied. Reference lists of included papers will be citation tracked for eligible studies using the Science Citation Index (Web of Science), as will relevant reviews of the literature identified through the searches.

Identification of grey literature

Unpublished literature will be identified through System for Information on Grey Literature in Europe (OpenGrey), ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science (Thomson). The authors of published conference proceedings will be contacted to obtain a full report of the findings where available. Data from conference proceedings will not be included in the review due to the limited information available for assessing inclusion, extracting data and undertaking the methodological quality assessment. There may also be differences in the data presented in conference proceedings and subsequent full study reports (28).

Inclusion / exclusion criteria

Types of instruments

Self-reported quantitative measures, such as surveys, checklists and questionnaires completed on paper or electronically are eligible for inclusion if they aim to measure one of the specified implementation outcomes. Questionnaires administered by interview or similar qualitative technique will be excluded.

Study design

Studies that aim to evaluate an implementation outcome instrument's measurement properties for use (or adaptation for use) in physical healthcare settings will be eligible for inclusion. Measurement properties include: reliability (internal consistency, test-retest

7

reliability and, if applicable, inter-rater reliability), validity (face and content validity, predictive and concurrent validity, convergent and discriminant validity), and dimensionality via the appropriate latent trait models (factor analysis, item response theory, item factor analysis, among others). Included studies can be published or unpublished full text original articles.

Setting and participants

This review will identify implementation outcome measures that have been developed for use in physical healthcare, grouped by different healthcare settings. Measures that have been developed for assessing implementation of interventions specifically for mental health conditions will be excluded as they have been identified in the existing Lewis et al review. However, in line with the review conducted by Lewis et al, we will include implementation outcomes instruments that are adaptable for use in physical healthcare settings. The eligibility of these generic instruments will be discussed with our stakeholder group. Implementation measures may target at any relevant stakeholder, such as organisation, provider, consumer/patient.

Types of implementation outcome measures

Quantitative instruments will be eligible for inclusion if they assess one of the implementation outcomes included in Proctor et al's taxonomy (7). To bring consistency and comparability to the field, Proctor et al conducted a review of the literature and proposed a working taxonomy of eight conceptual different, but interrelated, "implementation outcomes" that measure key elements of the implementation process, these are: feasibility, acceptability, appropriateness, adoption, penetration, fidelity, implementation cost and sustainability (7). For each outcome, they suggest the level of analysis (e.g. organisation, provider, consumer), theoretical basis (e.g. Rogers' theory of the diffusion of innovation (29)), overlapping constructs, salient implementation stage (e.g. early for adoption, ongoing for penetration, late for sustainability) and suitable research methods for measurement (e.g. survey, focus group, observation) (7).

These outcomes may be defined using different terms that describe the same underlying construct. The search terms include synonyms identified in the existing literature (see Table 2). Implementation outcomes may be measured at any implementation stage (e.g. pre-implementation, throughout implementation, post-implementation). Implementation outcomes may focus on attitudes, knowledge, behaviours, costs or number of participants receiving an intervention, etc.

In the Lewis et al review, measures of fidelity were eligible if they either 1) included assessments of implementation interventions or 2) were applicable to any evidence-based practice (i.e. not focussed on a specific practice (17), such as contingency management). This decision was made as measures of fidelity are extensively researched in specific treatment areas and tend to focus on specific interventions, thus limiting their generalisability to the field of implementation science. This review will exclude measures of fidelity on this basis.

Methodological quality of psychometric studies

Systematic reviews that investigate the measurement properties of instruments should assess: 1) the *methodological quality* of the psychometric studies and 2) the *psychometric quality* of the instrument and the appropriateness of statistical methods of evaluation, where is latter is dictated by the former (21). The *methodological quality* of the studies that investigate the measurement properties of the implementation instruments will be assessed using the COnsensus-Based Standards for the Selection of Health Status Measurement INstruments (COSMIN) quality criteria (21). The COSMIN checklist is a global measure of methodological quality, with separate criteria for nine different measurement properties. For each measurement property there are between 5-18 items used to assess the methodological quality of the study, each rated using a 4-point scale: "excellent", "good", "fair" or "poor". The lowest rating of any item for a particular measurement property is selected as the global score (21).

Psychometric quality of instruments and usability

We will use a structured checklist to evaluate the psychometric properties of the measures; this is currently under development and will be published on the Psychometrics and Measurement Lab website, at the Institute of Psychiatry, Psychology and Neuroscience at King's College London. This will cover: reliability (test-retest, internal consistency, inter-rater), validity (content, construct, and criterion validity) and dimensionality assessment (structural validity). The measures will be: 1) rated on whether the appropriate statistical methods were used, and 2) given a score based on results demonstrating good psychometric properties. The quality scores assigned to the results of each psychometric test will be based on published criteria and adjusted according to the identified studies, which will be used to set benchmarks for the field. This is in recognition that values will vary by field of study.

In the update of their systematic review of implementation outcomes in mental healthcare settings, Lewis et al are using a new measure of usability, which is currently under

9

development following a review of the literature and a consensus building exercise. The extent to which a measure is usable / pragmatic is an important aspect in this field, particularly where instruments are intended to be used as part of service evaluations (22). In applying the same tool as Lewis et al, we can compare findings between the mental and physical healthcare fields, thus contributing further to the implementation evidence-base.

Study screening

References identified by the search strategy will be entered into EndNote X8 bibliographic software and duplicates will be removed. Titles and abstracts will be screened independently by reviewers trained in systematic review methods and with experience of conducting psychometric research (ZK & LH). The full texts of all potentially relevant studies will be ordered and independently screened against the eligibility criteria in duplicate. Any discrepancies will be resolved by consensus with the wider research team and findings from the search will be presented in a PRISMA flow-chart (24,30).

Data extraction

Pre-designed extraction tables have been developed and piloted with studies included in the Lewis et al review (details below). Data will be entered into Microsoft Excel 2010 and checked for accuracy and completeness by a second reviewer. Authors will be contacted for missing data if necessary.

Instruments

For each of the seven implementation outcomes this review covers, the following data will be extracted for each instrument identified by the search strategy: authors and year of publication, country, name of instrument and version, number of items, construct and definition, setting, sample characteristics (e.g. gender, age, ethnicity), and characteristics of the intervention or innovation being implemented, level of analysis (i.e. organisation, provider, consumer), focus of measure (e.g. attitudes, knowledge, behaviour, or other) and implementation stage.

Psychometric studies

For each of the seven implementation outcomes, the following data will be extracted from the psychometric studies identified by the search strategy: authors and year of publication, country, name of instrument and version, type of psychometric study, setting, sample characteristics (e.g. gender, age, ethnicity), sample size, information needed to apply the COSMIN checklist and the results of the measurement properties. The reviewers will follow the comprehensive COSMIN manual on applying the methodological quality criteria to the

included studies (21). For each of the seven implementation outcomes, the methodological quality (COSMIN) ratings ("excellent", "good", "fair" or "poor") will be incorporated into tables including: authors and year of publication, name of instrument and type of measurement property assessed.

Data synthesis

Descriptive statistics will be used to present data on the number of instruments available and the number of measurement properties tested for each implementation outcome. A global score will be computed for: 1) methodological quality of psychometric studies and 2) psychometric quality of the instruments. These scores will be compared with those of the Lewis et al review (and review update). Due to the variability of instruments used in implementation research, quantitative evidence synthesis in the form of meta-analysis is deemed unfeasible (though this will be re-evaluated once the body of full-text original articles is in place).

DISCUSSION

Identifying implementation outcome measures and their measurement properties in wider healthcare settings is an important first step in informing the future research agenda in this field. It has been recommended that where instruments with promising measurement properties exist, priority should be given to further testing of these measures rather than developing new instruments (31). This review will identify priority areas where outcome instruments require further psychometric testing or where new measures are needed. In comparing the findings with previous reviews, we will have a better understanding of whether generic measures of implementation can be used, as opposed to context specific, with a view to standardising outcome measurement but not losing the salience of contextual factors.

The findings of this systematic review are intended to promote standardisation in the way implementation outcomes are measured, thus enabling comparison between studies, synthesis of findings in meta-analyses, and aiding the interpretation of research findings. It is important to note that implementation outcomes are amendable to both quantitative and qualitative methodologies. For example, *acceptability* can be explored using semi-structured interviews and focus groups to gain a more in-depth insight than a self-report questionnaire. Furthermore, other sources of quantitative data are useful, for example, routinely collected data can be used to measure *adoption*. The findings of this systematic review will inform mixed-method research projects, which blend the findings of quantitative and qualitative approaches (32).

11

Strengths and limitations

Systematic reviews of measurement properties are complex in terms of search strategies, methodological quality assessment and presentation of findings relating to the quality of the instruments. A validated search filter for identifying psychometric studies exists (33), however, for this review of implementation outcomes in all physical healthcare settings, our approach needed greater precision for screening to be manageable. One of the strengths of this review is its comprehensive search strategy, compared with previous reviews which tend to focus on a few broad terms and a particular setting. A further strength is the use of a methodological quality assessment tool, which to date, has not been applied to the research in this field. The COSMIN checklist was developed through an international Delphi exercise that sought consensus on standards for the design and statistical methods used in studies of measurement properties (21). We will also bespoke criteria for assessing the psychometric quality of the instruments, developed by the Psychometrics and Measurement Lab at King's College London, which will incorporate the suitability of the statistical method into the overall quality assessment of the instrument.

This review is limited to seven of the implementation outcomes proposed as part of Proctor et al's working taxonomy of implementation outcomes. Whilst these were identified by a search of the literature, they have not undergone consensus with key stakeholders and consumers to determine whether they constitute an exhaustive list. However, as Proctor et al acknowledge, these implementation outcomes constitute a working taxonomy and a strong starting point for measuring implementation outcomes across stakeholder level and implementation model, theory or framework.

Implications for research and practice

Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of the most suitable implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Word count = 3,270

12

Table 1. Search strings for Medline

1	translational medical research.sh.
2	diffusion of innovation.sh.
3	"implement*".ab,ti.
4	"adopt*".ab,ti.
5	"research utili* ".ab,ti.
6	"knowledge utili* ".ab,ti.
7	"knowledge mobil* ".ab,ti.
8	"knowledge transfer".ab,ti.
9	URE.ab,ti.
10	"use of research evidence".ab,ti.
11	"feasib*".ab,ti.
12	"acceptab*".ab,ti.
13	"appropriate*".ab,ti.
14	"adopt*".ab,ti.
15	"penetrat*".ab,ti.
16	"sustain*".ab,ti.
17	maintenance.ab,ti.
18	"transferab*".ab,ti.
19	"applicab*".ab,ti.
20	practicability.ab,ti.
21	"workab*".ab,ti.
22	uptake.ab,ti.
23	utility.ab,ti.
24	utilization.ab,ti.
25	utilisation.ab,ti.
26	credibility.ab,ti.
27	fit.ab,ti.
28	relevance.ab,ti.
29	"compatib*".ab,ti.
30	"suitab*".ab,ti.
31	usefulness.ab,ti.
32	reach.ab,ti.
33	spread.ab,ti.
34	coverage.ab,ti.
35	continuation.ab,ti.
36	"durab*".ab,ti.
37	"incorporat*".ab,ti.
38	"integrat*".ab,ti.
39	institutionalisation.ab,ti.
40	institutionalization.ab,ti.

41	routinization.ab,ti.
42	routinisation.ab,ti.
43	satisfaction.ab,ti.
44	agreeable.ab,ti.
45	discontinuation.ab,ti.
46	de-adoption.ab,ti.
47	normalisation.ab,ti.
48	normalization.ab,ti.
49	(implement* adj3 cost).ab,ti.
50	"internal consistency".ab,ti.
51	test-retest.ab,ti.
52	"test retest".ab,ti.
53	(reliability and (interrater or inter-rater or intrarater or intra-rater)).ab,ti.
54	"content validity".ab,ti.
55	"face validity".ab,ti.
56	"construct validity".ab,ti.
57	"criterion validity".ab,ti.
58	"structural validity".ab,ti.
59	"principal components analys*".ab,ti.
60	"factor analys* ".ab,ti.
61	"factor structure* ".ab,ti.
62	dimensionality.ab,ti.
63	"Item response model".ab,ti.
64	"Item response theory".ab,ti.
65	IRT.ab,ti.
66	MIMIC.ab,ti.
67	"classical test theory".ab,ti.
68	EFA.ab,ti.
69	CFA.ab,ti.
70	(exploratory or confirmatory).ab,ti.
71	factor.ab,ti.
72	70 and 71
73	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
74	11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49
75	50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 72

14

76	73 and 74 and 75
77	exp animals/ not humans.sh.
78	76 not 77

15

Table 2. Implementation outcomes and their synonyms

Acceptability	acceptab*[tiab] OR agreeable[tiab] OR satisfaction[tiab] OR credibility[tiab]
Adoption	adopt*[tiab] OR uptake[tiab] OR utility[tiab] OR utilization[tiab] OR utilisation[tiab] OR discontinuation[tiab] OR de-adoption[tiab]
Appropriateness	appropriate*[tiab] OR fit[tiab] OR relevance[tiab] OR compatib*[tiab] OR usefulness[tiab]
Feasibility	feasib*[tiab] OR suitab*[tiab] OR practicability[tiab] OR applicab*[tiab] OR workab*[tiab] OR transferab*[tiab]
Implementation cost	cost[tiab]
Penetration	penetrat*[tiab] OR reach[tiab] OR spread[tiab] OR coverage[tiab]
Sustainability	sustain*[tiab] OR maintenance[tiab] OR continuation[tiab] OR durab*[tiab] OR incorporat*[tiab] OR integrat*[tiab] OR institutionalisation[tiab] OR institutionalization[tiab] OR routinization[tiab] OR routinisation[tiab] OR normalisation[tiab] OR normalization[tiab]

16

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

NS is the Director of London Safety and Training Solutions Ltd, which provides quality and safety training and advisory services on a consultancy basis to healthcare organization globally. The other authors declare that they have no competing interests.

Authors' contributions

Zarnie Khadjesari (ZK) designed and drafted the protocol, and is guarantor of the review. Zarnie Khadjesari, Louise Hull (LH) and Nick Sevdalis (NS) conceived the study. ZK, LH and Silia Vitoratou (SV) piloted the data extraction forms. SV is developing the instrument quality criteria. All authors provided feedback on the review methods and contributed to the final manuscript.

Acknowledgements

17

18

REFERENCES

1. Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. Implement Sci. 2009;4:18.

2. Wilson PM, Sales A, Wensing M, Aarons GA, Flottorp S, Glidewell L, et al. Enhancing the reporting of implementation research. Implement Sci. 2017;12:13.

3. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation Hybrid Designs. Med Care. 2012 Mar;50(3):217–26.

4. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci. 2013;8:139.

5. Grimshaw J, Eccles M, Thomas R, MacLennan G, Ramsay C, Fraser C, et al. Toward evidence-based quality improvement. Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966-1998. J Gen Intern Med. 2006 Feb;21 Suppl 2:S14-20.

6. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation Research in Mental Health Services: an Emerging Science with Conceptual, Methodological, and Training challenges. Adm Policy Ment Health. 2009 Jan;36(1).

7. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011 Mar;38(2):65–76.

8. Powell BJ, Beidas RS, Lewis CC, Aarons GA, McMillen JC, Proctor EK, et al. Methods to Improve the Selection and Tailoring of Implementation Strategies. J Behav Health Serv Res. 2017 Apr 1;44(2):177–94.

9. Emmons KM, Weiner B, Fernandez ME, Tu S-P. Systems antecedents for dissemination and implementation: a review and analysis of measures. Health Educ Behav Off Publ Soc Public Health Educ. 2012 Feb;39(1):87–105.

10. Scott T, Mannion R, Davies H, Marshall M. The Quantitative Measurement of Organizational Culture in Health Care: A Review of the Available Instruments. Health Serv Res. 2003 Jun;38(3):923–45.

11. King T, Byers JF. A review of organizational culture instruments for nurse executives. J Nurs Adm. 2007 Jan;37(1):21–31.

12. Weiner BJ, Amick H, Lee S-YD. Review: Conceptualization and Measurement of Organizational Readiness for Change: A Review of the Literature in Health Services Research and Other Fields. Med Care Res Rev. 2008 Aug 1;65(4):379–436.

13. Squires JE, Estabrooks CA, Gustavsson P, Wallin L. Individual determinants of research utilization by nurses: a systematic review update. Implement Sci IS. 2011 Jan 5;6:1.

14. Chor KHB, Wisdom JP, Olin S-CS, Hoagwood KE, Horwitz SM. Measures for Predictors of Innovation Adoption. Adm Policy Ment Health Ment Health Serv Res. 2015 Sep 1;42(5):545–73.

19

15. Chaudoir SR, Dugan AG, Barr CHI. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. Implement Sci IS. 2013 Feb 17;8:22.

16. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci IS. 2009 Aug 7;4:50.

17. Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. Implement Sci IS. 2015 Nov 4;10:155.

18. Clinton-McHarg T, Yoong SL, Tzelepis F, Regan T, Fielding A, Skelton E, et al. Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: a systematic review. Implement Sci. 2016;11:148.

19. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci IS. 2015 Apr 21;10.

20. Terwee CB, Prinsen C a. C, Ricci Garotti MG, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2016 Apr;25(4):767–79.

21. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2012 May;21(4):651–7.

22. Glasgow RE, Riley WT. Pragmatic measures: what they are and why we need them. Am J Prev Med. 2013 Aug;45(2):237–43.

23. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349:g7647.

24. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009 Jul 21;339:b2535.

25. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2009 Apr;18(3):313–33.

26. Lewis CC, Stanick CF, Martinez RG, Weiner BJ, Kim M, Barwick M, et al. The Society for Implementation Research Collaboration Instrument Review Project: A methodology to promote rigorous evaluation. Implement Sci. 2015;10:2.

27. The Health Foundation. Evidence scan: Improvement Science. Lonon: The Health Foundation; 2011.

28. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2008.

29. Rogers E M. Diffusion of innovations. 4th edition. New York: The Free Press; 1995.

30. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009 Jul 21;339:b2700.

31. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. Implement Sci. 2014;9:118.

32. Aarons GA, Fettes DL, Sommerfeld DH, Palinkas L. Mixed Methods for Implementation Research: Application to Evidence-Based Practice Implementation and Staff Turnover in Community Based Organizations Providing Child Welfare Services. Child Maltreat. 2012 Feb;17(1).

33. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2009 Oct;18(8):1115–23.

BMJ Open

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

SCHOLARONE™
Manuscripts

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

Zarnie Khadjesari*, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, 16 De Crespigny Park, Camberwell, London SE5 8AF, United Kingdom
zarnie.khadjesari@kcl.ac.uk

Silia Vitoratou, Psychometrics and Measurement Lab, Biostatistics and Health Informatics Department, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
silia.vitoratou@kcl.ac.uk

Nick Sevdalis, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
nick.sevdalis@kcl.ac.uk

Louise Hull, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
louise.hull@kcl.ac.uk

*Corresponding author

1

ABSTRACT

Introduction

Over the past 10 years, research into methods that promote the uptake, implementation and sustainability of evidence-based interventions has gathered pace. Implementation outcomes, however, are defined in different ways and assessed by different measures, making it difficult to evaluate the effectiveness of implementation strategies and ultimately implementation success. The aim of this systematic review is to identify and appraise studies that assess the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings, to advance the use of precise and accurate measures.

Methods and analysis

The following databases will be searched for published literature from database inception to March 2017: MEDLINE, EMBASE, PsycINFO, CINAHL and the Cochrane Library. Grey literature will be sought via HMIC, OpenGrey, ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science. Reference lists of included studies and relevant reviews will be hand searched. Three search strings will be combined to identify eligible studies: 1) implementation literature, 2) implementation outcomes, 3) measurement properties. Screening of titles, abstracts and full papers will be assessed for eligibility by two reviewers independently and any discrepancies resolved via consensus with the wider research team. The methodological quality of the studies will be assessed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. A set of bespoke criteria to determine the quality of the instruments will be used, and the relationship between instrument usability and quality will be explored.

Dissemination

Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of the most suitable implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Systematic review registration

International Prospective Register of Systematic Reviews (PROSPERO): CRD42017065348

Key words: Systematic review; protocol; implementation outcomes; measurement properties; psychometric properties, implementation science

2

Strengths of this study

- We have designed a comprehensive search strategy for published and unpublished literature, and include a string of search terms for the type of measurement property.
- This will be the first systematic review of implementation outcomes that assesses the methodological quality of included studies.

Limitations of this study

- Due to the breadth of the setting (i.e. all physical healthcare settings), a validated search filter for measurement properties was not suitable as our approach needed greater precision for screening to be manageable.
- We selected a taxonomy of implementation outcomes to guide the selection of implementation outcomes in this review, however, there are several other models, theories and frameworks that could have guided the identification of measures in this field.

3

BACKGROUND

Routinely delivered, evidence-based practice is a principal objective of healthcare systems across the world. However, the so called 'evidence-to-practice gap' means it can take many years before patients benefit from evidence-based interventions, if at all, and when implementation is attempted it is often fraught with barriers (1). Over the past 10 years, research into methods that promote the uptake of evidence-based practices (i.e. implementation research) has substantially increased (2). However, due to the emerging state of the field and the breadth of disciplines it covers, implementation outcomes are defined in different ways and assessed by a variety of different measures, making it difficult to evaluate and compare the effectiveness of different implementation strategies- "methods or techniques used to enhance the adoption, implementation, and sustainability of a clinical programme or practice" (3–5). Implementation outcomes reflect the impact of efforts to implement evidence-based treatments, practices and services and are distinct from service and client/patient outcomes, which are essential but not sufficient for understanding implementation success, or failure (6). As such, it has been argued that implementation outcomes should be defined and measured in all studies of implementation (7). It has been proposed that implementation outcomes serve three functions: 1) indicate *implementation success*, which is a prerequisite for the effectiveness of treatment and quality of care approaches; 2) constitute proximal indicators of implementation processes; 3) provide important intermediate outcomes for service and client/patient outcomes (7). Accurate and precise measurement of implementation outcomes is thus vital for developing the evidence-base on effective implementation strategies (8).

Previous reviews have focussed on measures of system level antecedents to implementation (9), organisational level culture and readiness to change (10–12), and individual level determinants of research utilisation (13), as well as predictors of innovation adoption (14). Chaudoir et al identified 61 instruments that predict implementation of evidence-based interventions at multiple levels, with the majority assessing organisation, provider, and innovation-level constructs, as opposed to structural or patient-level constructs (15). More recently, reviews have taken a broader approach and identified instruments that assess the 37 constructs contained in the Consolidated Framework of Implementation Research (CFIR)- a meta-theoretical framework that aims to understand and/or explain influences on implementation outcomes (16–18). Furthermore, a review has focussed on identifying quantitative measures of the eight implementation outcomes included in Proctor et al's working taxonomy (17). Lewis et al identified 104 instruments that measure these constructs in mental healthcare settings: the vast majority of the instruments measured acceptability (n=50), followed by adoption (n=19), feasibility (n=8), cost (n=8), sustainability

4

(n=8), appropriateness (n=7) and penetration (n=4). The review highlighted that implementation outcome instrumentation is underdeveloped with regards to the number of instruments available and the measurement quality of instruments.

This systematic review will use Proctor et al's working taxonomy of implementation outcomes to guide the identification of implementation outcome instruments used in physical healthcare settings (i.e. excluding instruments specific to mental healthcare settings). The working taxonomy of implementation outcomes is relevant across stakeholder levels, stages of implementation, and can applied to different implementation models, theories and frameworks (19). This review will complement and allow direct comparison of the review of Lewis et al, whose review used the taxonomy to identify instruments used in mental health settings (17), where instruments were largely found to be specific to a particular intervention, behaviour and/or setting, to provide a complete picture of all available measures and their properties.

A review of systematic reviews of measurement properties of health-related outcome measurement instruments found that a number of them lacked comprehensive search strategies and methodological quality assessment. These are fundamental components of systematic review methodology, i.e. identifying all relevant literature in a field and providing information on the extent to which study results may be biased (20). The review identified 102 systematic reviews in a one-year period, and found only 59% had searched EMBASE (where searching MEDLINE and EMBASE databases is considered a minimal requirement by the authors (20)), 54% did not include search terms for measurement properties, and only 41% assessed the methodological quality of the studies (20).

This systematic review will address the methodological limitations of earlier reviews, namely, it will use a comprehensive search strategy, and it will assess the methodological quality of the included studies using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (21), which in-turn will inform the assessment of the instruments quality. In using a similar methodological approach to the Lewis et al review, we can compare our findings with those from the mental health field in terms of the methodological quality of the studies (the COSMIN will be applied to an update of the mental health review), the psychometric quality of the instruments for each outcome, and the impact of usability on the psychometric quality of the instruments- where pragmatic/usable measures are vital for the implementation of the instruments themselves (22). The purpose of this review is to promote and advance the use of precise and accurate measures of implementation outcomes across all physical healthcare settings.

5

METHODS

This review protocol has followed the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P) 2015 checklist (23,24). Amendments to the protocol are not anticipated, but will be reported in the publication of the results, should they occur.

Aim

- To evaluate the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings.

Objectives

- To systematically identify studies that assess the measurement properties of quantitative implementation outcome instruments
- To critically appraise the methodological quality of the evidence on measurement properties of implementation outcome measures using the COSMIN checklist
- To apply a bespoke criteria to determine the psychometric quality of the instruments
- To explore the relationship between instrument usability and quality.

Stakeholder group

This protocol has been developed with the support of an international stakeholder group, whose role is to ensure the research conducted by the Centre for Implementation Science, King's College London (where the review team are based) is of direct relevance to stakeholders' needs. The group consists of healthcare professionals, managers and academics working in the field of implementation science including journal editors and grant panel members. We have also received feedback on the protocol from the Centre for Implementation Science and King's Improvement Science research teams.

Search strategy

Three sets of search terms will be combined to identify studies that assess the measurement properties of instruments that measure implementation outcomes. The search strings describe: 1) the population / field of interest (i.e. implementation literature), 2) the constructs being measured (e.g. adoption) and 3) the measurement properties of instruments (e.g. test-retest reliability) (25). The first string of terms will be used to identify the implementation literature (such as implement* OR knowledge transfer), incorporating terms used by Lewis et al (26), the UK Health Foundation's scoping review on the concept and practice of improvement science (27), and index terms (e.g. MeSH) applied to Lewis et al's published

6

systematic review protocol (26) and publication of findings (17). The second string of terms will consist of the implementation outcomes included in Proctor et al's taxonomy and their synonyms (7,26). The third string of terms will relate to specific measurement properties of the instruments (such as internal consistency and content validity) (see Table 1).

We reviewed these search terms with our stakeholder groups to ensure they included all relevant synonyms. We will also conduct a supplementary search for the names of the instruments which are identified as eligible for inclusion in the review.

Published literature search

The following electronic databases will be searched using the search terms outlined above: MEDLINE, EMBASE, PsycINFO and HMIC (Health management Information Consortium) via the Ovid interface; CINAHL via the EBSCO Host interface; and the Cochrane library. Databases will be searched from inception to March 2017, there will be no language restrictions, a filter for studies in humans will be applied. Reference lists of included papers will be citation tracked for eligible studies using the Science Citation Index (Web of Science), as will relevant reviews of the literature identified through the searches.

Identification of grey literature

Unpublished literature will be identified through System for Information on Grey Literature in Europe (OpenGrey), ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science (Thomson). The authors of published conference proceedings will be contacted to obtain a full report of the findings where available. Data from conference proceedings will not be included in the review due to the limited information available for assessing inclusion, extracting data and undertaking the methodological quality assessment. There may also be differences in the data presented in conference proceedings and subsequent full study reports (28).

Inclusion / exclusion criteria

Types of instruments

Eligible measurement instruments are those designed to include indicator variables according to psychometric theory, as opposed to clinimetric scales (classification according to Fayers and Hand, 2002) (29). Psychometric scales consist of items which "do not alter or influence the underlying concept: they are merely aspects of it, or indicators of its magnitude" (29) p236, whereas clinimetric scales consist of items which are "merely constructing an index […] and need not to be indicator variables for the concept in question" (29) p237. These instruments may consist of surveys, checklists and/or questionnaires,

7

which can either be self-administered or administered by an interviewer or a rater and completed on paper or electronically.

Study design

Studies that aim to evaluate an implementation outcome instrument's measurement properties for use (or adaptation for use) in physical healthcare settings will be eligible for inclusion. Measurement properties include: reliability (internal consistency, test-retest reliability and, if applicable, inter-rater reliability), validity (face and content validity, predictive and concurrent validity, convergent and discriminant validity), and dimensionality via the appropriate latent trait models (factor analysis, item response theory, item factor analysis, among others). Included studies can be published or unpublished full text original articles, dissertations and theses.

Setting and participants

This review will identify implementation outcome measures that have been developed for use in physical healthcare, grouped by different healthcare settings. Measures that have been developed for assessing implementation of interventions specifically for mental health conditions will be excluded as they have been identified in the existing Lewis et al review. However, in line with the review conducted by Lewis et al, we will include implementation outcomes instruments that are adaptable for use in physical healthcare settings. The eligibility of these generic instruments will be discussed with our stakeholder group. Implementation measures may target at any relevant stakeholder, such as organisation, provider, consumer/patient.

Types of implementation outcome measures

Quantitative instruments will be eligible for inclusion if they assess one of the implementation outcomes included in Proctor et al's taxonomy (7). To bring consistency and comparability to the field, Proctor et al conducted a review of the literature and proposed a working taxonomy of eight conceptual different, but interrelated, "implementation outcomes" that measure key elements of the implementation process, these are: feasibility, acceptability, appropriateness, adoption, penetration, fidelity, implementation cost and sustainability (7). For each outcome, they suggest the level of analysis (e.g. organisation, provider, consumer), theoretical basis (e.g. Rogers' theory of the diffusion of innovation (30)), overlapping constructs, salient implementation stage (e.g. early for adoption, ongoing for penetration, late for sustainability) and suitable research methods for measurement (e.g. survey, focus group, observation) (7).

8

These outcomes may be defined using different terms that describe the same underlying construct. The search terms include synonyms identified in the existing literature (see Table 2). Implementation outcomes may be measured at any implementation stage (e.g. pre-implementation, throughout implementation, post-implementation). Implementation outcomes may focus on attitudes, knowledge, behaviours, costs or number of participants receiving an intervention, etc.

In the Lewis et al review, measures of fidelity were eligible if they either 1) included assessments of implementation interventions or 2) were applicable to any evidence-based practice (i.e. not focussed on a specific practice (17), such as contingency management). This decision was made as measures of fidelity are extensively researched in specific treatment areas and tend to focus on specific interventions, thus limiting their generalisability to the field of implementation science. This review will exclude measures of fidelity on this basis.

Methodological quality of psychometric studies

Systematic reviews that investigate the measurement properties of instruments should assess: 1) the *methodological quality* of the psychometric studies and 2) the *psychometric quality* of the instrument and the appropriateness of statistical methods of evaluation, where is latter is dictated by the former (21). The *methodological quality* of the studies that investigate the measurement properties of the implementation instruments will be assessed using the COnsensus-Based Standards for the Selection of Health Status Measurement INstruments (COSMIN) quality criteria (21). The COSMIN checklist is a global measure of methodological quality, with separate criteria for nine different measurement properties. For each measurement property there are between 5-18 items used to assess the methodological quality of the study, each rated using a 4-point scale: "excellent", "good", "fair" or "poor". The lowest rating of any item for a particular measurement property is selected as the global score (21).

Psychometric quality of instruments and usability

We will use a structured checklist to evaluate the psychometric properties of the measures; this is currently under development and will be published on the Psychometrics and Measurement Lab website, at the Institute of Psychiatry, Psychology and Neuroscience at King's College London. This will cover: reliability (test-retest, internal consistency, inter-rater), validity (content, construct, and criterion validity) and dimensionality assessment (structural validity). The measures will be: 1) rated on whether the appropriate statistical methods were used, and 2) given a score based on results demonstrating good

9

psychometric properties. The quality scores assigned to the results of each psychometric test will be based on published criteria and adjusted according to the identified studies, which will be used to set benchmarks for the field. This is in recognition that values will vary by field of study.

In the update of their systematic review of implementation outcomes in mental healthcare settings, Lewis et al are using a new measure of usability, which is currently under development following a review of the literature and a consensus building exercise. The extent to which a measure is usable / pragmatic is an important aspect in this field, particularly where instruments are intended to be used as part of service evaluations (22). In applying the same tool as Lewis et al, we can compare findings between the mental and physical healthcare fields, thus contributing further to the implementation evidence-base.

Study screening

References identified by the search strategy will be entered into EndNote X8 bibliographic software and duplicates will be removed. Titles and abstracts will be screened independently by reviewers trained in systematic review methods and with experience of conducting psychometric research. The full texts of all potentially relevant studies will be ordered and independently screened against the eligibility criteria in duplicate. Any discrepancies will be resolved by consensus with the wider research team and findings from the search will be presented in a PRISMA flow-chart (24,31).

Data extraction

Pre-designed extraction tables have been developed and piloted with studies included in the Lewis et al review (details below). Data will be entered into Microsoft Excel 2010 and checked for accuracy and completeness by a second reviewer. Authors will be contacted for missing data if necessary.

Instruments

For each of the seven implementation outcome instruments this review identifies, the following data will be extracted for each instrument identified by the search strategy: authors and year of publication, country, name of instrument and version, number of items, construct and definition, setting, sample characteristics (e.g. gender, age, ethnicity), and characteristics of the intervention or innovation being implemented, level of analysis (i.e. organisation, provider, consumer), focus of measure (e.g. attitudes, knowledge, behaviour, or other) and implementation stage (e.g. pre-implementation, throughout implementation, post-implementation).

10

Psychometric studies

For each of the seven implementation outcomes, the following data will be extracted from the psychometric studies identified by the search strategy: authors and year of publication, country, name of instrument and version, type of psychometric study, setting, sample characteristics (e.g. gender, age, ethnicity), sample size, information needed to apply the COSMIN checklist and the results of the measurement properties. The reviewers will follow the comprehensive COSMIN manual on applying the methodological quality criteria to the included studies (21). For each of the seven implementation outcomes, the methodological quality (COSMIN) ratings ("excellent", "good", "fair" or "poor") will be incorporated into tables including: authors and year of publication, name of instrument, type of measurement property assessed and information needed to assess usability.

Data synthesis

Descriptive statistics will be used to present data on the number of instruments available and the number of measurement properties tested for each implementation outcome. A global score will be computed for: 1) methodological quality of psychometric studies and 2) psychometric quality of the instruments. The instrument quality scores will be included in tables similar to those presented in the review conducted by Lewis et al (17), which includes the number and percentage of instruments with a rating of 1 or more for each outcome and a table of summary statistics of instrument quality ratings by outcome. The average quality rating for each measurement property for each outcome will also be presented graphically. The COSMIN ratings, the instrument quality ratings and the usability scores will be compared with those of the Lewis et al review (and review update). Due to the variability of instruments used in implementation research, quantitative evidence synthesis in the form of meta-analysis is deemed unfeasible (though this will be re-evaluated once the body of full-text original articles is in place).

DISCUSSION

Identifying implementation outcome measures and their measurement properties in wider healthcare settings is an important first step in informing the future research agenda in this field. It has been recommended that where instruments with promising measurement properties exist, priority should be given to further testing of these measures rather than developing new instruments (32). This review will identify priority areas where implementation outcome instruments require further psychometric testing or where new measures are needed. In comparing the findings with previous reviews, we will have a better understanding of whether generic measures of implementation outcomes can be

11

used, as opposed to context specific, with a view to standardising implementation outcome measurement but not losing the salience of contextual factors.

The findings of this systematic review are intended to promote standardisation in the way implementation outcomes are measured, thus enabling comparison between studies, synthesis of findings in meta-analyses, and aiding the interpretation of research findings. It is important to note that implementation outcomes are amendable to both quantitative and qualitative methodologies. For example, *acceptability* can be explored using semi-structured interviews and focus groups to gain a more in-depth insight than a self-report questionnaire. Furthermore, other sources of quantitative data are useful, for example, routinely collected data can be used to measure *adoption*. The findings of this systematic review will inform mixed-method research projects, which blend the findings of quantitative and qualitative approaches (33).

Strengths and limitations

Systematic reviews of measurement properties are complex in terms of search strategies, methodological quality assessment and presentation of findings relating to the quality of the instruments. A validated search filter for identifying psychometric studies exists (34), however, for this review of implementation outcomes in all physical healthcare settings, our approach needed greater precision for screening to be manageable. One of the strengths of this review is its comprehensive search strategy, compared with previous reviews which tend to focus on a few broad terms and a particular setting. A further strength is the use of a methodological quality assessment tool, which to date, has not been applied to the research in this field. The COSMIN checklist was developed through an international Delphi exercise that sought consensus on standards for the design and statistical methods used in studies of measurement properties (21). We will also use bespoke criteria for assessing the psychometric quality of the instruments, developed by the Psychometrics and Measurement Lab at King's College London, which will incorporate the suitability of the statistical method into the overall quality assessment of the instrument.

This review is limited to seven of the implementation outcomes proposed as part of Proctor et al's working taxonomy of implementation outcomes. Whilst these were identified by a search of the literature, they have not undergone consensus with key stakeholders and consumers to determine whether they constitute an exhaustive list. However, as Proctor et al acknowledge, these implementation outcomes constitute a working taxonomy and a strong starting point for measuring implementation outcomes across stakeholder level and implementation model, theory or framework.

12

Implications for research and practice

Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of the most suitable implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Word count = 3,466

Table 1. Search strings for Medline

1	translational medical research.sh.
2	diffusion of innovation.sh.
3	"implement*".ab,ti.
4	"adopt*".ab,ti.
5	"research utili* ".ab,ti.
6	"knowledge utili* ".ab,ti.
7	"knowledge mobil* ".ab,ti.
8	"knowledge transfer".ab,ti.
9	URE.ab,ti.
10	"use of research evidence".ab,ti.
11	"feasib*".ab,ti.
12	"acceptab*".ab,ti.
13	"appropriate*".ab,ti.
14	"adopt*".ab,ti.
15	"penetrat*".ab,ti.
16	"sustain*".ab,ti.
17	maintenance.ab,ti.
18	"transferab*".ab,ti.
19	"applicab*".ab,ti.
20	practicability.ab,ti.
21	"workab*".ab,ti.
22	uptake.ab,ti.
23	utility.ab,ti.
24	utilization.ab,ti.
25	utilisation.ab,ti.
26	credibility.ab,ti.
27	fit.ab,ti.
28	relevance.ab,ti.
29	"compatib*".ab,ti.
30	"suitab*".ab,ti.
31	usefulness.ab,ti.
32	reach.ab,ti.
33	spread.ab,ti.
34	coverage.ab,ti.
35	continuation.ab,ti.
36	"durab*".ab,ti.
37	"incorporat*".ab,ti.
38	"integrat*".ab,ti.
39	institutionalisation.ab,ti.
40	institutionalization.ab,ti.

14

41	routinization.ab,ti.
42	routinisation.ab,ti.
43	satisfaction.ab,ti.
44	agreeable.ab,ti.
45	discontinuation.ab,ti.
46	de-adoption.ab,ti.
47	normalisation.ab,ti.
48	normalization.ab,ti.
49	(implement* adj3 cost).ab,ti.
50	"internal consistency".ab,ti.
51	test-retest.ab,ti.
52	"test retest".ab,ti.
53	(reliability and (interrater or inter-rater or intrarater or intra-rater)).ab,ti.
54	"content validity".ab,ti.
55	"face validity".ab,ti.
56	"construct validity".ab,ti.
57	"criterion validity".ab,ti.
58	"structural validity".ab,ti.
59	"concurrent validity".ab,ti.
60	"predictive validity".ab,ti.
61	"convergent validity".ab,ti.
62	"discriminant validity".ab,ti.
63	"principal components analys*".ab,ti.
64	"factor analys* ".ab,ti.
65	"factor structure* ".ab,ti.
66	dimensionality.ab,ti.
67	"Item response model".ab,ti.
68	"Item response theory".ab,ti.
69	IRT.ab,ti.
70	MIMIC.ab,ti.
71	"classical test theory".ab,ti.
72	EFA.ab,ti.
73	CFA.ab,ti.
74	(exploratory or confirmatory).ab,ti.
75	factor.ab,ti.
76	74 and 75
77	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
78	11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49

15

79	50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 76
80	77 and 78 and 79
81	exp animals/ not humans.sh.
82	80 not 81

16

Table 2. Implementation outcomes and their synonyms

Acceptability	acceptab*, agreeable, satisfaction, credibility
Adoption	adopt*, uptake, utility, utilization, utilisation, discontinuation, de-adoption
Appropriateness	appropriate*, fit, relevance, compatib*, usefulness
Feasibility	feasib*, suitab*, practicability, applicab*, workab*, transferab*
Implementation cost	Cost
Penetration	penetrat*, reach, spread, coverage
Sustainability	sustain*, maintenance, continuation, durab*, incorporat*, integrat*, institutionalisation, institutionalization, routinization, routinisation, normalisation, normalization

17

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

NS is the Director of London Safety and Training Solutions Ltd, which provides quality and safety training and advisory services on a consultancy basis to healthcare organization globally. The other authors declare that they have no competing interests.

Authors' contributions

Zarnie Khadjesari (ZK) designed and drafted the protocol, and is guarantor of the review. Zarnie Khadjesari, Louise Hull (LH) and Nick Sevdalis (NS) conceived the study. ZK, LH and Silia Vitoratou (SV) piloted the data extraction forms. SV is developing the instrument quality criteria. All authors provided feedback on the review methods and contributed to the final manuscript.

Acknowledgements

19

REFERENCES

1. Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. Implement Sci. 2009;4:18.

2. Wilson PM, Sales A, Wensing M, Aarons GA, Flottorp S, Glidewell L, et al. Enhancing the reporting of implementation research. Implement Sci. 2017;12:13.

3. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation Hybrid Designs. Med Care. 2012 Mar;50(3):217–26.

4. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci. 2013;8:139.

5. Grimshaw J, Eccles M, Thomas R, MacLennan G, Ramsay C, Fraser C, et al. Toward evidence-based quality improvement. Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966-1998. J Gen Intern Med. 2006 Feb;21 Suppl 2:S14-20.

6. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation Research in Mental Health Services: an Emerging Science with Conceptual, Methodological, and Training challenges. Adm Policy Ment Health. 2009 Jan;36(1).

7. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011 Mar;38(2):65–76.

8. Powell BJ, Beidas RS, Lewis CC, Aarons GA, McMillen JC, Proctor EK, et al. Methods to Improve the Selection and Tailoring of Implementation Strategies. J Behav Health Serv Res. 2017 Apr 1;44(2):177–94.

9. Emmons KM, Weiner B, Fernandez ME, Tu S-P. Systems antecedents for dissemination and implementation: a review and analysis of measures. Health Educ Behav Off Publ Soc Public Health Educ. 2012 Feb;39(1):87–105.

10. Scott T, Mannion R, Davies H, Marshall M. The Quantitative Measurement of Organizational Culture in Health Care: A Review of the Available Instruments. Health Serv Res. 2003 Jun;38(3):923–45.

11. King T, Byers JF. A review of organizational culture instruments for nurse executives. J Nurs Adm. 2007 Jan;37(1):21–31.

12. Weiner BJ, Amick H, Lee S-YD. Review: Conceptualization and Measurement of Organizational Readiness for Change: A Review of the Literature in Health Services Research and Other Fields. Med Care Res Rev. 2008 Aug 1;65(4):379–436.

13. Squires JE, Estabrooks CA, Gustavsson P, Wallin L. Individual determinants of research utilization by nurses: a systematic review update. Implement Sci IS. 2011 Jan 5;6:1.

14. Chor KHB, Wisdom JP, Olin S-CS, Hoagwood KE, Horwitz SM. Measures for Predictors of Innovation Adoption. Adm Policy Ment Health Ment Health Serv Res. 2015 Sep 1;42(5):545–73.

20

15. Chaudoir SR, Dugan AG, Barr CHI. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. Implement Sci IS. 2013 Feb 17;8:22.

16. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci IS. 2009 Aug 7;4:50.

17. Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. Implement Sci IS. 2015 Nov 4;10:155.

18. Clinton-McHarg T, Yoong SL, Tzelepis F, Regan T, Fielding A, Skelton E, et al. Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: a systematic review. Implement Sci. 2016;11:148.

19. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci IS. 2015 Apr 21;10.

20. Terwee CB, Prinsen C a. C, Ricci Garotti MG, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2016 Apr;25(4):767–79.

21. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2012 May;21(4):651–7.

22. Glasgow RE, Riley WT. Pragmatic measures: what they are and why we need them. Am J Prev Med. 2013 Aug;45(2):237–43.

23. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;350:g7647.

24. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009 Jul 21;339:b2535.

25. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2009 Apr;18(3):313–33.

26. Lewis CC, Stanick CF, Martinez RG, Weiner BJ, Kim M, Barwick M, et al. The Society for Implementation Research Collaboration Instrument Review Project: A methodology to promote rigorous evaluation. Implement Sci. 2015;10:2.

27. The Health Foundation. Evidence scan: Improvement Science. London: The Health Foundation; 2011.

28. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2008.

21

29. Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. J R Stat Soc Ser A Stat Soc. 2002 Jun 1;165(2):233–53.

30. Rogers E M. Diffusion of innovations. 4th edition. New York: The Free Press; 1995.

31. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009 Jul 21;339:b2700.

32. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. Implement Sci. 2014;9:118.

33. Aarons GA, Fettes DL, Sommerfeld DH, Palinkas L. Mixed Methods for Implementation Research: Application to Evidence-Based Practice Implementation and Staff Turnover in Community Based Organizations Providing Child Welfare Services. Child Maltreat. 2012 Feb;17(1).

34. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2009 Oct;18(8):1115–23.

22

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Manuscript page number
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	N/A
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	18
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	6
Support:			
Sources	5a	Indicate sources of financial or other support for the review	18
Sponsor	5b	Provide name for the review funder and/or sponsor	N/A
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	18
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-5
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	6
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	7-9
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	7

Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	6-7, 14-16
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	10
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	10
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	10-11
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	10-11
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	11
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	9
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	N/A
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	N/A
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	N/A
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	11
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	N/A
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	N/A

* It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.

BMJ Open

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

SCHOLARONE™
Manuscripts

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

Zarnie Khadjesari*, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, 16 De Crespigny Park, Camberwell, London SE5 8AF, United Kingdom
zarnie.khadjesari@kcl.ac.uk

Silia Vitoratou, Psychometrics and Measurement Lab, Biostatistics and Health Informatics Department, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
silia.vitoratou@kcl.ac.uk

Nick Sevdalis, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
nick.sevdalis@kcl.ac.uk

Louise Hull, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
louise.hull@kcl.ac.uk

*Corresponding author

1

ABSTRACT

Introduction

Over the past 10 years, research into methods that promote the uptake, implementation and sustainability of evidence-based interventions has gathered pace. Implementation outcomes, however, are defined in different ways and assessed by different measures, making it difficult to evaluate the effectiveness of implementation strategies and ultimately implementation success. The aim of this systematic review is to identify and appraise studies that assess the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings, to advance the use of precise and accurate measures.

Methods and analysis

The following databases will be searched for published literature from database inception to March 2017: MEDLINE, EMBASE, PsycINFO, CINAHL and the Cochrane Library. Grey literature will be sought via HMIC, OpenGrey, ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science. Reference lists of included studies and relevant reviews will be hand searched. Three search strings will be combined to identify eligible studies: 1) implementation literature, 2) implementation outcomes, 3) measurement properties. Screening of titles, abstracts and full papers will be assessed for eligibility by two reviewers independently and any discrepancies resolved via consensus with the wider research team. The methodological quality of the studies will be assessed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. A set of bespoke criteria to determine the quality of the instruments will be used, and the relationship between instrument usability and quality will be explored.

Dissemination

Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of the most suitable implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Systematic review registration

International Prospective Register of Systematic Reviews (PROSPERO): CRD42017065348

2

Strengths of this study

- We have designed a comprehensive search strategy for published and unpublished literature, and include a string of search terms for the type of measurement property.
- This will be the first systematic review of implementation outcomes that assesses the methodological quality of included studies.

Limitations of this study

- Due to the breadth of the setting (i.e. all physical healthcare settings), a validated search filter for measurement properties was not suitable as our approach needed greater precision for screening to be manageable.
- We selected a taxonomy of implementation outcomes to guide the selection of implementation outcomes in this review, however, there are several other models, theories and frameworks that could have guided the identification of measures in this field.

3

BACKGROUND

Routinely delivered, evidence-based practice is a principal objective of healthcare systems across the world. However, the so called 'evidence-to-practice gap' means it can take many years before patients benefit from evidence-based interventions, if at all, and when implementation is attempted it is often fraught with barriers (1). Over the past 10 years, research into methods that promote the uptake of evidence-based practices (i.e. implementation research) has substantially increased (2). However, due to the emerging state of the field and the breadth of disciplines it covers, implementation outcomes are defined in different ways and assessed by a variety of different measures, making it difficult to evaluate and compare the effectiveness of different implementation strategies- "methods or techniques used to enhance the adoption, implementation, and sustainability of a clinical programme or practice" (3–5). Implementation outcomes reflect the impact of efforts to implement evidence-based treatments, practices and services and are distinct from service and client/patient outcomes, which are essential but not sufficient for understanding implementation success, or failure (6). As such, it has been argued that implementation outcomes should be defined and measured in all studies of implementation (7). It has been proposed that implementation outcomes serve three functions: 1) indicate *implementation success*, which is a prerequisite for the effectiveness of treatment and quality of care approaches; 2) constitute proximal indicators of implementation processes; 3) provide important intermediate outcomes for service and client/patient outcomes (7). Accurate and precise measurement of implementation outcomes is thus vital for developing the evidence-base on effective implementation strategies (8).

Previous reviews have focussed on measures of system level antecedents to implementation (9), organisational level culture and readiness to change (10–12), and individual level determinants of research utilisation (13), as well as predictors of innovation adoption (14). Chaudoir et al identified 61 instruments that predict implementation of evidence-based interventions at multiple levels, with the majority assessing organisation, provider, and innovation-level constructs, as opposed to structural or patient-level constructs (15). More recently, reviews have taken a broader approach and identified instruments that assess the 37 constructs contained in the Consolidated Framework of Implementation Research (CFIR)- a meta-theoretical framework that aims to understand and/or explain influences on implementation outcomes (16–18). Furthermore, a review has focussed on identifying quantitative measures of the eight implementation outcomes included in Proctor et al's working taxonomy (17). Lewis et al identified 104 instruments that measure these constructs in mental healthcare settings: the vast majority of the instruments measured acceptability (n=50), followed by adoption (n=19), feasibility (n=8), cost (n=8), sustainability

(n=8), appropriateness (n=7) and penetration (n=4). The review highlighted that implementation outcome instrumentation is underdeveloped with regards to the number of instruments available and the measurement quality of instruments.

This systematic review will use Proctor et al's working taxonomy of implementation outcomes to guide the identification of implementation outcome instruments used in physical healthcare settings (i.e. excluding instruments specific to mental healthcare settings). The working taxonomy of implementation outcomes is relevant across stakeholder levels, stages of implementation, and can applied to different implementation models, theories and frameworks (19). This review will complement and allow direct comparison of the review of Lewis et al, whose review used the taxonomy to identify instruments used in mental health settings (17), where instruments were largely found to be specific to a particular intervention, behaviour and/or setting, to provide a complete picture of all available measures and their properties.

A review of systematic reviews of measurement properties of health-related outcome measurement instruments found that a number of them lacked comprehensive search strategies and methodological quality assessment. These are fundamental components of systematic review methodology, i.e. identifying all relevant literature in a field and providing information on the extent to which study results may be biased (20). The review identified 102 systematic reviews in a one-year period, and found only 59% had searched EMBASE (where searching MEDLINE and EMBASE databases is considered a minimal requirement by the authors (20)), 54% did not include search terms for measurement properties, and only 41% assessed the methodological quality of the studies (20).

This systematic review will address the methodological limitations of earlier reviews, namely, it will use a comprehensive search strategy, and it will assess the methodological quality of the included studies using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (21), which in-turn will inform the assessment of the instruments quality. In using a similar methodological approach to the Lewis et al review, we can compare our findings with those from the mental health field in terms of the methodological quality of the studies (the COSMIN will be applied to an update of the mental health review), the psychometric quality of the instruments for each outcome, and the impact of usability on the psychometric quality of the instruments- where pragmatic/usable measures are vital for the implementation of the instruments themselves (22). The purpose of this review is to promote and advance the use of precise and accurate measures of implementation outcomes across all physical healthcare settings.

METHODS

This review protocol has followed the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P) 2015 checklist (23,24). Amendments to the protocol are not anticipated, but will be reported in the publication of the results, should they occur.

Aim

- To evaluate the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings.

Objectives

- To systematically identify studies that assess the measurement properties of quantitative implementation outcome instruments
- To critically appraise the methodological quality of the evidence on measurement properties of implementation outcome measures using the COSMIN checklist
- To apply a bespoke criteria to determine the psychometric quality of the instruments
- To explore the relationship between instrument usability and quality.

Stakeholder group

This protocol has been developed with the support of an international stakeholder group, whose role is to ensure the research conducted by the Centre for Implementation Science, King's College London (where the review team are based) is of direct relevance to stakeholders' needs. The group consists of healthcare professionals, managers and academics working in the field of implementation science including journal editors and grant panel members. We have also received feedback on the protocol from the Centre for Implementation Science and King's Improvement Science research teams.

Search strategy

Three sets of search terms will be combined to identify studies that assess the measurement properties of instruments that measure implementation outcomes. The search strings describe: 1) the population / field of interest (i.e. implementation literature), 2) the constructs being measured (e.g. adoption) and 3) the measurement properties of instruments (e.g. test-retest reliability) (25). The first string of terms will be used to identify the implementation literature (such as implement* OR knowledge transfer), incorporating terms used by Lewis et al (26), the UK Health Foundation's scoping review on the concept and practice of improvement science (27), and index terms (e.g. MeSH) applied to Lewis et al's published

6

systematic review protocol (26) and publication of findings (17). The second string of terms will consist of the implementation outcomes included in Proctor et al's taxonomy and their synonyms (7,26). The third string of terms will relate to specific measurement properties of the instruments (such as internal consistency and content validity) (see Table 1).

We reviewed these search terms with our stakeholder groups to ensure they included all relevant synonyms. We will also conduct a supplementary search for the names of the instruments which are identified as eligible for inclusion in the review.

Published literature search

The following electronic databases will be searched using the search terms outlined above: MEDLINE, EMBASE, PsycINFO and HMIC (Health management Information Consortium) via the Ovid interface; CINAHL via the EBSCO Host interface; and the Cochrane library. Databases will be searched from inception to March 2017, there will be no language restrictions, a filter for studies in humans will be applied. Reference lists of included papers will be citation tracked for eligible studies using the Science Citation Index (Web of Science), as will relevant reviews of the literature identified through the searches.

Identification of grey literature

Unpublished literature will be identified through System for Information on Grey Literature in Europe (OpenGrey), ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science (Thomson). The authors of published conference proceedings will be contacted to obtain a full report of the findings where available. Data from conference proceedings will not be included in the review due to the limited information available for assessing inclusion, extracting data and undertaking the methodological quality assessment. There may also be differences in the data presented in conference proceedings and subsequent full study reports (28).

Inclusion / exclusion criteria

Types of instruments

Eligible measurement instruments are those designed to include indicator variables according to psychometric theory, as opposed to clinimetric scales (classification according to Fayers and Hand, 2002) (29). Psychometric scales consist of items which "do not alter or influence the underlying concept: they are merely aspects of it, or indicators of its magnitude" (29) p236, whereas clinimetric scales consist of items which are "merely constructing an index […] and need not to be indicator variables for the concept in question" (29) p237. These instruments may consist of surveys, checklists and/or questionnaires,

7

which can either be self-administered or administered by an interviewer or a rater and completed on paper or electronically.

Study design

Studies that aim to evaluate an implementation outcome instrument's measurement properties for use (or adaptation for use) in physical healthcare settings will be eligible for inclusion. Measurement properties include: reliability (internal consistency, test-retest reliability and, if applicable, inter-rater reliability), validity (face and content validity, predictive and concurrent validity, convergent and discriminant validity), and dimensionality via the appropriate latent trait models (factor analysis, item response theory, item factor analysis, among others). Included studies can be published or unpublished full text original articles, dissertations and theses.

Setting and participants

This review will identify implementation outcome measures that have been developed for use in physical healthcare, grouped by different healthcare settings. Measures that have been developed for assessing implementation of interventions specifically for mental health conditions will be excluded as they have been identified in the existing Lewis et al review. However, in line with the review conducted by Lewis et al, we will include implementation outcomes instruments that are adaptable for use in physical healthcare settings. The eligibility of these generic instruments will be discussed with our stakeholder group. Implementation measures may target at any relevant stakeholder, such as organisation, provider, consumer/patient.

Types of implementation outcome measures

Quantitative instruments will be eligible for inclusion if they assess one of the implementation outcomes included in Proctor et al's taxonomy (7). To bring consistency and comparability to the field, Proctor et al conducted a review of the literature and proposed a working taxonomy of eight conceptual different, but interrelated, "implementation outcomes" that measure key elements of the implementation process, these are: feasibility, acceptability, appropriateness, adoption, penetration, fidelity, implementation cost and sustainability (7). For each outcome, they suggest the level of analysis (e.g. organisation, provider, consumer), theoretical basis (e.g. Rogers' theory of the diffusion of innovation (30)), overlapping constructs, salient implementation stage (e.g. early for adoption, ongoing for penetration, late for sustainability) and suitable research methods for measurement (e.g. survey, focus group, observation) (7).

These outcomes may be defined using different terms that describe the same underlying construct. The search terms include synonyms identified in the existing literature (see Table 2). Implementation outcomes may be measured at any implementation stage (e.g. pre-implementation, throughout implementation, post-implementation). Implementation outcomes may focus on attitudes, knowledge, behaviours, costs or number of participants receiving an intervention, etc.

In the Lewis et al review, measures of fidelity were eligible if they either 1) included assessments of implementation interventions or 2) were applicable to any evidence-based practice (i.e. not focussed on a specific practice (17), such as contingency management). This decision was made as measures of fidelity are extensively researched in specific treatment areas and tend to focus on specific interventions, thus limiting their generalisability to the field of implementation science. This review will exclude measures of fidelity on this basis.

Methodological quality of psychometric studies

Systematic reviews that investigate the measurement properties of instruments should assess: 1) the *methodological quality* of the psychometric studies and 2) the *psychometric quality* of the instrument and the appropriateness of statistical methods of evaluation, where is latter is dictated by the former (21). The *methodological quality* of the studies that investigate the measurement properties of the implementation instruments will be assessed using the COnsensus-Based Standards for the Selection of Health Status Measurement INstruments (COSMIN) quality criteria (21). The COSMIN checklist is a global measure of methodological quality, with separate criteria for nine different measurement properties. For each measurement property there are between 5-18 items used to assess the methodological quality of the study, each rated using a 4-point scale: "excellent", "good", "fair" or "poor". The lowest rating of any item for a particular measurement property is selected as the global score (21).

Psychometric quality of instruments and usability

We will use a structured checklist to evaluate the psychometric properties of the measures; this is currently under development and will be published on the Psychometrics and Measurement Lab website, at the Institute of Psychiatry, Psychology and Neuroscience at King's College London. This will cover: reliability (test-retest, internal consistency, inter-rater), validity (content, construct, and criterion validity) and dimensionality assessment (structural validity). The measures will be: 1) rated on whether the appropriate statistical methods were used, and 2) given a score based on results demonstrating good

9

psychometric properties. The quality scores assigned to the results of each psychometric test will be based on published criteria and adjusted according to the identified studies, which will be used to set benchmarks for the field. This is in recognition that values will vary by field of study.

In the update of their systematic review of implementation outcomes in mental healthcare settings, Lewis et al are using a new measure of usability, which is currently under development following a review of the literature and a consensus building exercise. The extent to which a measure is usable / pragmatic is an important aspect in this field, particularly where instruments are intended to be used as part of service evaluations (22). In applying the same tool as Lewis et al, we can compare findings between the mental and physical healthcare fields, thus contributing further to the implementation evidence-base.

Study screening

References identified by the search strategy will be entered into EndNote X8 bibliographic software and duplicates will be removed. Titles and abstracts will be screened independently by reviewers trained in systematic review methods and with experience of conducting psychometric research. The full texts of all potentially relevant studies will be ordered and independently screened against the eligibility criteria in duplicate. Any discrepancies will be resolved by consensus with the wider research team and findings from the search will be presented in a PRISMA flow-chart (24,31).

Data extraction

Pre-designed extraction tables have been developed and piloted with studies included in the Lewis et al review (details below). Data will be entered into Microsoft Excel 2010 and checked for accuracy and completeness by a second reviewer. Authors will be contacted for missing data if necessary.

Instruments

For each of the seven implementation outcome instruments this review identifies, the following data will be extracted for each instrument identified by the search strategy: authors and year of publication, country, name of instrument and version, number of items, construct and definition, setting, sample characteristics (e.g. gender, age, ethnicity), and characteristics of the intervention or innovation being implemented, level of analysis (i.e. organisation, provider, consumer), focus of measure (e.g. attitudes, knowledge, behaviour, or other) and implementation stage (e.g. pre-implementation, throughout implementation, post-implementation).

10

Psychometric studies

For each of the seven implementation outcomes, the following data will be extracted from the psychometric studies identified by the search strategy: authors and year of publication, country, name of instrument and version, type of psychometric study, setting, sample characteristics (e.g. gender, age, ethnicity), sample size, information needed to apply the COSMIN checklist and the results of the measurement properties. The reviewers will follow the comprehensive COSMIN manual on applying the methodological quality criteria to the included studies (21). For each of the seven implementation outcomes, the methodological quality (COSMIN) ratings ("excellent", "good", "fair" or "poor") will be incorporated into tables including: authors and year of publication, name of instrument, type of measurement property assessed and information needed to assess usability.

Data synthesis

Descriptive statistics will be used to present data on the number of instruments available and the number of measurement properties tested for each implementation outcome. A global score will be computed for: 1) methodological quality of psychometric studies and 2) psychometric quality of the instruments. The instrument quality scores will be included in tables similar to those presented in the review conducted by Lewis et al (17), which includes the number and percentage of instruments with a rating of 1 or more for each outcome and a table of summary statistics of instrument quality ratings by outcome. The average quality rating for each measurement property for each outcome will also be presented graphically. The COSMIN ratings, the instrument quality ratings and the usability scores will be compared with those of the Lewis et al review (and review update). Due to the variability of instruments used in implementation research, quantitative evidence synthesis in the form of meta-analysis is deemed unfeasible (though this will be re-evaluated once the body of full-text original articles is in place).

DISCUSSION

Identifying implementation outcome measures and their measurement properties in wider healthcare settings is an important first step in informing the future research agenda in this field. It has been recommended that where instruments with promising measurement properties exist, priority should be given to further testing of these measures rather than developing new instruments (32). This review will identify priority areas where implementation outcome instruments require further psychometric testing or where new measures are needed. In comparing the findings with previous reviews, we will have a better understanding of whether generic measures of implementation outcomes can be

11

used, as opposed to context specific, with a view to standardising implementation outcome measurement but not losing the salience of contextual factors.

The findings of this systematic review are intended to promote standardisation in the way implementation outcomes are measured, thus enabling comparison between studies, synthesis of findings in meta-analyses, and aiding the interpretation of research findings. It is important to note that implementation outcomes are amenable to both quantitative and qualitative methodologies. For example, *acceptability* can be explored using semi-structured interviews and focus groups to gain a more in-depth insight than a self-report questionnaire. Furthermore, other sources of quantitative data are useful, for example, routinely collected data can be used to measure *adoption*. The findings of this systematic review will inform mixed-method research projects, which blend the findings of quantitative and qualitative approaches (33).

Strengths and limitations

Systematic reviews of measurement properties are complex in terms of search strategies, methodological quality assessment and presentation of findings relating to the quality of the instruments. A validated search filter for identifying psychometric studies exists (34), however, for this review of implementation outcomes in all physical healthcare settings, our approach needed greater precision for screening to be manageable. One of the strengths of this review is its comprehensive search strategy, compared with previous reviews which tend to focus on a few broad terms and a particular setting. A further strength is the use of a methodological quality assessment tool, which to date, has not been applied to the research in this field. The COSMIN checklist was developed through an international Delphi exercise that sought consensus on standards for the design and statistical methods used in studies of measurement properties (21). We will also use bespoke criteria for assessing the psychometric quality of the instruments, developed by the Psychometrics and Measurement Lab at King's College London, which will incorporate the suitability of the statistical method into the overall quality assessment of the instrument.

This review is limited to seven of the implementation outcomes proposed as part of Proctor et al's working taxonomy of implementation outcomes. Whilst these were identified by a search of the literature, they have not undergone consensus with key stakeholders and consumers to determine whether they constitute an exhaustive list. However, as Proctor et al acknowledge, these implementation outcomes constitute a working taxonomy and a strong starting point for measuring implementation outcomes across stakeholder level and implementation model, theory or framework.

12

Implications for research and practice

Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of the most suitable implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Word count = 3,466

13

Table 1. Search strings for Medline

1	translational medical research.sh.
2	diffusion of innovation.sh.
3	"implement*".ab,ti.
4	"adopt*".ab,ti.
5	"research utili* ".ab,ti.
6	"knowledge utili* ".ab,ti.
7	"knowledge mobil* ".ab,ti.
8	"knowledge transfer".ab,ti.
9	URE.ab,ti.
10	"use of research evidence".ab,ti.
11	"feasib*".ab,ti.
12	"acceptab*".ab,ti.
13	"appropriate*".ab,ti.
14	"adopt*".ab,ti.
15	"penetrat*".ab,ti.
16	"sustain*".ab,ti.
17	maintenance.ab,ti.
18	"transferab*".ab,ti.
19	"applicab*".ab,ti.
20	practicability.ab,ti.
21	"workab*".ab,ti.
22	uptake.ab,ti.
23	utility.ab,ti.
24	utilization.ab,ti.
25	utilisation.ab,ti.
26	credibility.ab,ti.
27	fit.ab,ti.
28	relevance.ab,ti.
29	"compatib*".ab,ti.
30	"suitab*".ab,ti.
31	usefulness.ab,ti.
32	reach.ab,ti.
33	spread.ab,ti.
34	coverage.ab,ti.
35	continuation.ab,ti.
36	"durab*".ab,ti.
37	"incorporat*".ab,ti.
38	"integrat*".ab,ti.
39	institutionalisation.ab,ti.
40	institutionalization.ab,ti.

14

41	routinization.ab,ti.
42	routinisation.ab,ti.
43	satisfaction.ab,ti.
44	agreeable.ab,ti.
45	discontinuation.ab,ti.
46	de-adoption.ab,ti.
47	normalisation.ab,ti.
48	normalization.ab,ti.
49	(implement* adj3 cost).ab,ti.
50	"internal consistency".ab,ti.
51	test-retest.ab,ti.
52	"test retest".ab,ti.
53	(reliability and (interrater or inter-rater or intrarater or intra-rater)).ab,ti.
54	"content validity".ab,ti.
55	"face validity".ab,ti.
56	"construct validity".ab,ti.
57	"criterion validity".ab,ti.
58	"structural validity".ab,ti.
59	"concurrent validity".ab,ti.
60	"predictive validity".ab,ti.
61	"convergent validity".ab,ti.
62	"discriminant validity".ab,ti.
63	"principal components analys*".ab,ti.
64	"factor analys* ".ab,ti.
65	"factor structure* ".ab,ti.
66	dimensionality.ab,ti.
67	"Item response model".ab,ti.
68	"Item response theory".ab,ti.
69	IRT.ab,ti.
70	MIMIC.ab,ti.
71	"classical test theory".ab,ti.
72	EFA.ab,ti.
73	CFA.ab,ti.
74	(exploratory or confirmatory).ab,ti.
75	factor.ab,ti.
76	74 and 75
77	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
78	11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49

15

79	50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 76
80	77 and 78 and 79
81	exp animals/ not humans.sh.
82	80 not 81

16

Table 2. Implementation outcomes and their synonyms

Acceptability	acceptab*, agreeable, satisfaction, credibility
Adoption	adopt*, uptake, utility, utilization, utilisation, discontinuation, de-adoption
Appropriateness	appropriate*, fit, relevance, compatib*, usefulness
Feasibility	feasib*, suitab*, practicability, applicab*, workab*, transferab*
Implementation cost	Cost
Penetration	penetrat*, reach, spread, coverage
Sustainability	sustain*, maintenance, continuation, durab*, incorporat*, integrat*, institutionalisation, institutionalization, routinization, routinisation, normalisation, normalization

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

NS is the Director of London Safety and Training Solutions Ltd, which provides quality and safety training and advisory services on a consultancy basis to healthcare organization globally. The other authors declare that they have no competing interests.

Authors' contributions

Zarnie Khadjesari (ZK) designed and drafted the protocol, and is guarantor of the review. Zarnie Khadjesari, Louise Hull (LH) and Nick Sevdalis (NS) conceived the study. ZK, LH and Silia Vitoratou (SV) piloted the data extraction forms. SV is developing the instrument quality criteria. All authors provided feedback on the review methods and contributed to the final manuscript.

Acknowledgements

19

REFERENCES

1. Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. Implement Sci. 2009;4:18.

2. Wilson PM, Sales A, Wensing M, Aarons GA, Flottorp S, Glidewell L, et al. Enhancing the reporting of implementation research. Implement Sci. 2017;12:13.

3. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation Hybrid Designs. Med Care. 2012;50(3):217–26.

4. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci. 2013;8:139.

5. Grimshaw J, Eccles M, Thomas R, MacLennan G, Ramsay C, Fraser C, et al. Toward evidence-based quality improvement. Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966-1998. J Gen Intern Med. 2006;21 Suppl 2:S14-20.

6. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation Research in Mental Health Services: an Emerging Science with Conceptual, Methodological, and Training challenges. Adm Policy Ment Health. 2009;36(1).

7. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011;38(2):65–76.

8. Powell BJ, Beidas RS, Lewis CC, Aarons GA, McMillen JC, Proctor EK, et al. Methods to Improve the Selection and Tailoring of Implementation Strategies. J Behav Health Serv Res. 2017;44(2):177–94.

9. Emmons KM, Weiner B, Fernandez ME, Tu S-P. Systems antecedents for dissemination and implementation: a review and analysis of measures. Health Educ Behav. 2012;39(1):87–105.

10. Scott T, Mannion R, Davies H, Marshall M. The Quantitative Measurement of Organizational Culture in Health Care: A Review of the Available Instruments. Health Serv Res. 2003;38(3):923–45.

11. King T, Byers JF. A review of organizational culture instruments for nurse executives. J Nurs Adm. 2007;37(1):21–31.

12. Weiner BJ, Amick H, Lee S-YD. Review: Conceptualization and Measurement of Organizational Readiness for Change: A Review of the Literature in Health Services Research and Other Fields. Med Care Res Rev. 2008;65(4):379–436.

13. Squires JE, Estabrooks CA, Gustavsson P, Wallin L. Individual determinants of research utilization by nurses: a systematic review update. Implement Sci. 2011;6:1.

14. Chor KHB, Wisdom JP, Olin S-CS, Hoagwood KE, Horwitz SM. Measures for Predictors of Innovation Adoption. Adm Policy Ment Health Ment Health Serv Res. 2015;42(5):545–73.

20

15. Chaudoir SR, Dugan AG, Barr CHI. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. Implement Sci. 2013;8:22.

16. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci. 2009;4:50.

17. Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. Implement Sci. 2015;10:155.

18. Clinton-McHarg T, Yoong SL, Tzelepis F, Regan T, Fielding A, Skelton E, et al. Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: a systematic review. Implement Sci. 2016;11:148.

19. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci. 2015;10.

20. Terwee CB, Prinsen C a. C, Ricci Garotti MG, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res. 2016;25(4):767–79.

21. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651–7.

22. Glasgow RE, Riley WT. Pragmatic measures: what they are and why we need them. Am J Prev Med. 2013;45(2):237–43.

23. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;350:g7647.

24. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.

25. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res. 2009;18(3):313–33.

26. Lewis CC, Stanick CF, Martinez RG, Weiner BJ, Kim M, Barwick M, et al. The Society for Implementation Research Collaboration Instrument Review Project: A methodology to promote rigorous evaluation. Implement Sci. 2015;10:2.

27. The Health Foundation. Evidence scan: Improvement Science. London: The Health Foundation; 2011.

28. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2008.

29. Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. J R Stat Soc Ser A. 2002;165(2):233–53.

21

30. Rogers E M. Diffusion of innovations. 4th edition. New York: The Free Press; 1995.

31. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.

32. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. Implement Sci. 2014;9:118.

33. Aarons GA, Fettes DL, Sommerfeld DH, Palinkas L. Mixed Methods for Implementation Research: Application to Evidence-Based Practice Implementation and Staff Turnover in Community Based Organizations Providing Child Welfare Services. Child Maltreat. 2012;17(1).

34. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115–23.

22

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Manuscript page number
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	N/A
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	18
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	6
Support:			
Sources	5a	Indicate sources of financial or other support for the review	18
Sponsor	5b	Provide name for the review funder and/or sponsor	N/A
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	18
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-5
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	6
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	7-9
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	7

Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	6-7, 14-16
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	10
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	10
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	10-11
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	10-11
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	11
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	9
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	N/A
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	N/A
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	N/A
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	11
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	N/A
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	N/A

* It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

BMJ Open

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

SCHOLARONE™
Manuscripts

Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol

Zarnie Khadjesari*, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, 16 De Crespigny Park, Camberwell, London SE5 8AF, United Kingdom
zarnie.khadjesari@kcl.ac.uk

Silia Vitoratou, Psychometrics and Measurement Lab, Biostatistics and Health Informatics Department, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
silia.vitoratou@kcl.ac.uk

Nick Sevdalis, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
nick.sevdalis@kcl.ac.uk

Louise Hull, Centre for Implementation Science, Department of Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, United Kingdom
louise.hull@kcl.ac.uk

*Corresponding author

ABSTRACT

Introduction

Over the past 10 years, research into methods that promote the uptake, implementation and sustainability of evidence-based interventions has gathered pace. However, implementation outcomes are defined in different ways and assessed by different measures- the extent to which these are valid and reliable is unknown. The aim of this systematic review is to identify and appraise studies that assess the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings, to advance the use of precise and accurate measures.

Methods and analysis

The following databases will be searched from inception to March 2017: MEDLINE, EMBASE, PsycINFO, CINAHL and the Cochrane Library. Grey literature will be sought via HMIC, OpenGrey, ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science. Reference lists of included studies and relevant reviews will be hand searched. Three search strings will be combined to identify eligible studies: 1) implementation literature, 2) implementation outcomes, 3) measurement properties. Screening of titles, abstracts and full papers will be assessed for eligibility by two reviewers independently and any discrepancies resolved via consensus with the wider team. The methodological quality of the studies will be assessed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. A set of bespoke criteria to determine the quality of the instruments will be used, and the relationship between instrument usability and quality will be explored.

Ethics and Dissemination

Ethical approval is not necessary for systematic review protocols. Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Systematic review registration

International Prospective Register of Systematic Reviews (PROSPERO): CRD42017065348

Key words: Systematic review; protocol; implementation outcomes; measurement properties; psychometric properties, implementation science

2

Strengths of this study

- We have designed a comprehensive search strategy for published and unpublished literature, and include a string of search terms for the type of measurement property.
- This will be the first systematic review of implementation outcomes that assesses the methodological quality of included studies.

Limitations of this study

- Due to the breadth of the setting (i.e. all physical healthcare settings), a validated search filter for measurement properties was not suitable as our approach needed greater precision for screening to be manageable.
- We selected a taxonomy of implementation outcomes to guide the selection of implementation outcomes in this review, however, there are several other models, theories and frameworks that could have guided the identification of measures in this field.

3

BACKGROUND

Routinely delivered, evidence-based practice is a principal objective of healthcare systems across the world. However, the so called 'evidence-to-practice gap' means it can take many years before patients benefit from evidence-based interventions, if at all, and when implementation is attempted it is often fraught with barriers (1). Over the past 10 years, research into methods that promote the uptake of evidence-based practices (i.e. implementation research) has substantially increased (2). However, due to the emerging state of the field and the breadth of disciplines it covers, implementation outcomes are defined in different ways and assessed by a variety of different measures, making it difficult to evaluate and compare the effectiveness of different implementation strategies- "methods or techniques used to enhance the adoption, implementation, and sustainability of a clinical programme or practice" (3–5). Implementation outcomes reflect the impact of efforts to implement evidence-based treatments, practices and services and are distinct from service and client/patient outcomes, which are essential but not sufficient for understanding implementation success, or failure (6). As such, it has been argued that implementation outcomes should be defined and measured in all studies of implementation (7). It has been proposed that implementation outcomes serve three functions: 1) indicate *implementation success*, which is a prerequisite for the effectiveness of treatment and quality of care approaches; 2) constitute proximal indicators of implementation processes; 3) provide important intermediate outcomes for service and client/patient outcomes (7). Accurate and precise measurement of implementation outcomes is thus vital for developing the evidence-base on effective implementation strategies (8).

Previous reviews have focussed on measures of system level antecedents to implementation (9), organisational level culture and readiness to change (10–12), and individual level determinants of research utilisation (13), as well as predictors of innovation adoption (14). Chaudoir et al identified 61 instruments that predict implementation of evidence-based interventions at multiple levels, with the majority assessing organisation, provider, and innovation-level constructs, as opposed to structural or patient-level constructs (15). More recently, reviews have taken a broader approach and identified instruments that assess the 37 constructs contained in the Consolidated Framework of Implementation Research (CFIR)- a meta-theoretical framework that aims to understand and/or explain influences on implementation outcomes (16–18). Furthermore, a review has focussed on identifying quantitative measures of the eight implementation outcomes included in Proctor et al's working taxonomy (17). Lewis et al identified 104 instruments that measure these constructs in mental healthcare settings: the vast majority of the instruments measured acceptability (n=50), followed by adoption (n=19), feasibility (n=8), cost (n=8), sustainability

4

(n=8), appropriateness (n=7) and penetration (n=4). The review highlighted that implementation outcome instrumentation is underdeveloped with regards to the number of instruments available and the measurement quality of instruments.

This systematic review will use Proctor et al's working taxonomy of implementation outcomes to guide the identification of implementation outcome instruments used in physical healthcare settings (i.e. excluding instruments specific to mental healthcare settings). The working taxonomy of implementation outcomes is relevant across stakeholder levels, stages of implementation, and can applied to different implementation models, theories and frameworks (19). This review will complement and allow direct comparison of the review of Lewis et al, whose review used the taxonomy to identify instruments used in mental health settings (17), where instruments were largely found to be specific to a particular intervention, behaviour and/or setting, to provide a complete picture of all available measures and their properties.

A review of systematic reviews of measurement properties of health-related outcome measurement instruments found that a number of them lacked comprehensive search strategies and methodological quality assessment. These are fundamental components of systematic review methodology, i.e. identifying all relevant literature in a field and providing information on the extent to which study results may be biased (20). The review identified 102 systematic reviews in a one-year period, and found only 59% had searched EMBASE (where searching MEDLINE and EMBASE databases is considered a minimal requirement by the authors (20)), 54% did not include search terms for measurement properties, and only 41% assessed the methodological quality of the studies (20).

This systematic review will address the methodological limitations of earlier reviews, namely, it will use a comprehensive search strategy, and it will assess the methodological quality of the included studies using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (21), which in-turn will inform the assessment of the instruments quality. In using a similar methodological approach to the Lewis et al review, we can compare our findings with those from the mental health field in terms of the methodological quality of the studies (the COSMIN will be applied to an update of the mental health review), the psychometric quality of the instruments for each outcome, and the impact of usability on the psychometric quality of the instruments- where pragmatic/usable measures are vital for the implementation of the instruments themselves (22). The purpose of this review is to promote and advance the use of precise and accurate measures of implementation outcomes across all physical healthcare settings.

5

METHODS

This review protocol has followed the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P) 2015 checklist (23,24). Amendments to the protocol are not anticipated, but will be reported in the publication of the results, should they occur.

Aim

- To evaluate the measurement properties of quantitative implementation outcome instruments used in physical healthcare settings.

Objectives

- To systematically identify studies that assess the measurement properties of quantitative implementation outcome instruments
- To critically appraise the methodological quality of the evidence on measurement properties of implementation outcome measures using the COSMIN checklist
- To apply a bespoke criteria to determine the psychometric quality of the instruments
- To explore the relationship between instrument usability and quality.

Stakeholder group

This protocol has been developed with the support of an international stakeholder group, whose role is to ensure the research conducted by the Centre for Implementation Science, King's College London (where the review team are based) is of direct relevance to stakeholders' needs. The group consists of healthcare professionals, managers and academics working in the field of implementation science including journal editors and grant panel members. We have also received feedback on the protocol from the Centre for Implementation Science and King's Improvement Science research teams.

Search strategy

Three sets of search terms will be combined to identify studies that assess the measurement properties of instruments that measure implementation outcomes. The search strings describe: 1) the population / field of interest (i.e. implementation literature), 2) the constructs being measured (e.g. adoption) and 3) the measurement properties of instruments (e.g. test-retest reliability) (25). The first string of terms will be used to identify the implementation literature (such as implement* OR knowledge transfer), incorporating terms used by Lewis et al (26), the UK Health Foundation's scoping review on the concept and practice of improvement science (27), and index terms (e.g. MeSH) applied to Lewis et al's published

6

systematic review protocol (26) and publication of findings (17). The second string of terms will consist of the implementation outcomes included in Proctor et al's taxonomy and their synonyms (7,26). The third string of terms will relate to specific measurement properties of the instruments (such as internal consistency and content validity) (see Table 1).

We reviewed these search terms with our stakeholder groups to ensure they included all relevant synonyms. We will also conduct a supplementary search for the names of the instruments which are identified as eligible for inclusion in the review.

Published literature search

The following electronic databases will be searched using the search terms outlined above: MEDLINE, EMBASE, PsycINFO and HMIC (Health management Information Consortium) via the Ovid interface; CINAHL via the EBSCO Host interface; and the Cochrane library. Databases will be searched from inception to March 2017, there will be no language restrictions, a filter for studies in humans will be applied. Reference lists of included papers will be citation tracked for eligible studies using the Science Citation Index (Web of Science), as will relevant reviews of the literature identified through the searches.

Identification of grey literature

Unpublished literature will be identified through System for Information on Grey Literature in Europe (OpenGrey), ProQuest for theses and Web of Science Conference Proceedings Citation Index-Science (Thomson). The authors of published conference proceedings will be contacted to obtain a full report of the findings where available. Data from conference proceedings will not be included in the review due to the limited information available for assessing inclusion, extracting data and undertaking the methodological quality assessment. There may also be differences in the data presented in conference proceedings and subsequent full study reports (28).

Inclusion / exclusion criteria

Types of instruments

Eligible measurement instruments are those designed to include indicator variables according to psychometric theory, as opposed to clinimetric scales (classification according to Fayers and Hand, 2002) (29). Psychometric scales consist of items which "do not alter or influence the underlying concept: they are merely aspects of it, or indicators of its magnitude" (29) p236, whereas clinimetric scales consist of items which are "merely constructing an index […] and need not to be indicator variables for the concept in question" (29) p237. These instruments may consist of surveys, checklists and/or questionnaires,

7

which can either be self-administered or administered by an interviewer or a rater and completed on paper or electronically.

Study design

Studies that aim to evaluate an implementation outcome instrument's measurement properties for use (or adaptation for use) in physical healthcare settings will be eligible for inclusion. Measurement properties include: reliability (internal consistency, test-retest reliability and, if applicable, inter-rater reliability), validity (face and content validity, predictive and concurrent validity, convergent and discriminant validity), and dimensionality via the appropriate latent trait models (factor analysis, item response theory, item factor analysis, among others). Included studies can be published or unpublished full text original articles, dissertations and theses.

Setting and participants

This review will identify implementation outcome measures that have been developed for use in physical healthcare, grouped by different healthcare settings. Measures that have been developed for assessing implementation of interventions specifically for mental health conditions will be excluded as they have been identified in the existing Lewis et al review. However, in line with the review conducted by Lewis et al, we will include implementation outcomes instruments that are adaptable for use in physical healthcare settings. The eligibility of these generic instruments will be discussed with our stakeholder group. Implementation measures may target at any relevant stakeholder, such as organisation, provider, consumer/patient.

Types of implementation outcome measures

Quantitative instruments will be eligible for inclusion if they assess one of the implementation outcomes included in Proctor et al's taxonomy (7). To bring consistency and comparability to the field, Proctor et al conducted a review of the literature and proposed a working taxonomy of eight conceptual different, but interrelated, "implementation outcomes" that measure key elements of the implementation process, these are: feasibility, acceptability, appropriateness, adoption, penetration, fidelity, implementation cost and sustainability (7). For each outcome, they suggest the level of analysis (e.g. organisation, provider, consumer), theoretical basis (e.g. Rogers' theory of the diffusion of innovation (30)), overlapping constructs, salient implementation stage (e.g. early for adoption, ongoing for penetration, late for sustainability) and suitable research methods for measurement (e.g. survey, focus group, observation) (7).

8

These outcomes may be defined using different terms that describe the same underlying construct. The search terms include synonyms identified in the existing literature (see Table 2). Implementation outcomes may be measured at any implementation stage (e.g. pre-implementation, throughout implementation, post-implementation). Implementation outcomes may focus on attitudes, knowledge, behaviours, costs or number of participants receiving an intervention, etc.

In the Lewis et al review, measures of fidelity were eligible if they either 1) included assessments of implementation interventions or 2) were applicable to any evidence-based practice (i.e. not focussed on a specific practice (17), such as contingency management). This decision was made as measures of fidelity are extensively researched in specific treatment areas and tend to focus on specific interventions, thus limiting their generalisability to the field of implementation science. This review will exclude measures of fidelity on this basis.

Methodological quality of psychometric studies

Systematic reviews that investigate the measurement properties of instruments should assess: 1) the *methodological quality* of the psychometric studies and 2) the *psychometric quality* of the instrument and the appropriateness of statistical methods of evaluation, where is latter is dictated by the former (21). The *methodological quality* of the studies that investigate the measurement properties of the implementation instruments will be assessed using the COnsensus-Based Standards for the Selection of Health Status Measurement INstruments (COSMIN) quality criteria (21). The COSMIN checklist is a global measure of methodological quality, with separate criteria for nine different measurement properties. For each measurement property there are between 5-18 items used to assess the methodological quality of the study, each rated using a 4-point scale: "excellent", "good", "fair" or "poor". The lowest rating of any item for a particular measurement property is selected as the global score (21).

Psychometric quality of instruments and usability

We will use a structured checklist to evaluate the psychometric properties of the measures; this is currently under development and will be published on the Psychometrics and Measurement Lab website, at the Institute of Psychiatry, Psychology and Neuroscience at King's College London. This will cover: reliability (test-retest, internal consistency, inter-rater), validity (content, construct, and criterion validity) and dimensionality assessment (structural validity). The measures will be: 1) rated on whether the appropriate statistical methods were used, and 2) given a score based on results demonstrating good

9

psychometric properties. The quality scores assigned to the results of each psychometric test will be based on published criteria and adjusted according to the identified studies, which will be used to set benchmarks for the field. This is in recognition that values will vary by field of study.

In the update of their systematic review of implementation outcomes in mental healthcare settings, Lewis et al are using a new measure of usability, which is currently under development following a review of the literature and a consensus building exercise. The extent to which a measure is usable / pragmatic is an important aspect in this field, particularly where instruments are intended to be used as part of service evaluations (22). In applying the same tool as Lewis et al, we can compare findings between the mental and physical healthcare fields, thus contributing further to the implementation evidence-base.

Study screening

References identified by the search strategy will be entered into EndNote X8 bibliographic software and duplicates will be removed. Titles and abstracts will be screened independently by reviewers trained in systematic review methods and with experience of conducting psychometric research. The full texts of all potentially relevant studies will be ordered and independently screened against the eligibility criteria in duplicate. Any discrepancies will be resolved by consensus with the wider research team and findings from the search will be presented in a PRISMA flow-chart (24,31).

Data extraction

Pre-designed extraction tables have been developed and piloted with studies included in the Lewis et al review (details below). Data will be entered into Microsoft Excel 2010 and checked for accuracy and completeness by a second reviewer. Authors will be contacted for missing data if necessary.

Instruments

For each of the seven implementation outcome instruments this review identifies, the following data will be extracted for each instrument identified by the search strategy: authors and year of publication, country, name of instrument and version, number of items, construct and definition, setting, sample characteristics (e.g. gender, age, ethnicity), and characteristics of the intervention or innovation being implemented, level of analysis (i.e. organisation, provider, consumer), focus of measure (e.g. attitudes, knowledge, behaviour, or other) and implementation stage (e.g. pre-implementation, throughout implementation, post-implementation).

10

Psychometric studies

For each of the seven implementation outcomes, the following data will be extracted from the psychometric studies identified by the search strategy: authors and year of publication, country, name of instrument and version, type of psychometric study, setting, sample characteristics (e.g. gender, age, ethnicity), sample size, information needed to apply the COSMIN checklist and the results of the measurement properties. The reviewers will follow the comprehensive COSMIN manual on applying the methodological quality criteria to the included studies (21). For each of the seven implementation outcomes, the methodological quality (COSMIN) ratings ("excellent", "good", "fair" or "poor") will be incorporated into tables including: authors and year of publication, name of instrument, type of measurement property assessed and information needed to assess usability.

Data synthesis

Descriptive statistics will be used to present data on the number of instruments available and the number of measurement properties tested for each implementation outcome. A global score will be computed for: 1) methodological quality of psychometric studies and 2) psychometric quality of the instruments. The instrument quality scores will be included in tables similar to those presented in the review conducted by Lewis et al (17), which includes the number and percentage of instruments with a rating of 1 or more for each outcome and a table of summary statistics of instrument quality ratings by outcome. The average quality rating for each measurement property for each outcome will also be presented graphically. The COSMIN ratings, the instrument quality ratings and the usability scores will be compared with those of the Lewis et al review (and review update). Due to the variability of instruments used in implementation research, quantitative evidence synthesis in the form of meta-analysis is deemed unfeasible (though this will be re-evaluated once the body of full-text original articles is in place).

DISCUSSION

Identifying implementation outcome measures and their measurement properties in wider healthcare settings is an important first step in informing the future research agenda in this field. It has been recommended that where instruments with promising measurement properties exist, priority should be given to further testing of these measures rather than developing new instruments (32). This review will identify priority areas where implementation outcome instruments require further psychometric testing or where new measures are needed. In comparing the findings with previous reviews, we will have a better understanding of whether generic measures of implementation outcomes can be

11

used, as opposed to context specific, with a view to standardising implementation outcome measurement but not losing the salience of contextual factors.

The findings of this systematic review are intended to promote standardisation in the way implementation outcomes are measured, thus enabling comparison between studies, synthesis of findings in meta-analyses, and aiding the interpretation of research findings. It is important to note that implementation outcomes are amenable to both quantitative and qualitative methodologies. For example, *acceptability* can be explored using semi-structured interviews and focus groups to gain a more in-depth insight than a self-report questionnaire. Furthermore, other sources of quantitative data are useful, for example, routinely collected data can be used to measure *adoption*. The findings of this systematic review will inform mixed-method research projects, which blend the findings of quantitative and qualitative approaches (33).

Strengths and limitations

Systematic reviews of measurement properties are complex in terms of search strategies, methodological quality assessment and presentation of findings relating to the quality of the instruments. A validated search filter for identifying psychometric studies exists (34), however, for this review of implementation outcomes in all physical healthcare settings, our approach needed greater precision for screening to be manageable. One of the strengths of this review is its comprehensive search strategy, compared with previous reviews which tend to focus on a few broad terms and a particular setting. A further strength is the use of a methodological quality assessment tool, which to date, has not been applied to the research in this field. The COSMIN checklist was developed through an international Delphi exercise that sought consensus on standards for the design and statistical methods used in studies of measurement properties (21). We will also use bespoke criteria for assessing the psychometric quality of the instruments, developed by the Psychometrics and Measurement Lab at King's College London, which will incorporate the suitability of the statistical method into the overall quality assessment of the instrument.

This review is limited to seven of the implementation outcomes proposed as part of Proctor et al's working taxonomy of implementation outcomes. Whilst these were identified by a search of the literature, they have not undergone consensus with key stakeholders and consumers to determine whether they constitute an exhaustive list. However, as Proctor et al acknowledge, these implementation outcomes constitute a working taxonomy and a strong starting point for measuring implementation outcomes across stakeholder level and implementation model, theory or framework.

12

Ethics and Dissemination

This systematic review will identify, appraise and synthesise secondary data found in published and unpublished studies, therefore ethical approval is not necessary.

Findings of the review will be published in an open access peer reviewed journal and presented at international conferences, such as the Society for Implementation Research Collaboration (SIRC). The findings will also be disseminated to healthcare professionals, managers, patients, the public and policy makers via the Centre for Implementation Science and King's Improvement Science websites, reported in their newsletters, integrated into resources and guides provided by these centres, and tweeted by the Collaboration for Leadership in Applied Health Research and Care South London (@CLAHRC_SL). Researchers and healthcare professionals can use the findings of this systematic review to guide the selection of the most suitable implementation outcomes instruments, based on their psychometric quality, to assess the impact of their implementation efforts. The findings will also provide a useful guide for reviewers of papers and grants to determine the psychometric quality of the measures used in implementation research.

Word count = 3,569

Table 1. Search strings for Medline

1	translational medical research.sh.
2	diffusion of innovation.sh.
3	"implement*".ab,ti.
4	"adopt*".ab,ti.
5	"research utili* ".ab,ti.
6	"knowledge utili* ".ab,ti.
7	"knowledge mobil* ".ab,ti.
8	"knowledge transfer".ab,ti.
9	URE.ab,ti.
10	"use of research evidence".ab,ti.
11	"feasib*".ab,ti.
12	"acceptab*".ab,ti.
13	"appropriate*".ab,ti.
14	"adopt*".ab,ti.
15	"penetrat*".ab,ti.
16	"sustain*".ab,ti.
17	maintenance.ab,ti.
18	"transferab*".ab,ti.
19	"applicab*".ab,ti.
20	practicability.ab,ti.
21	"workab*".ab,ti.
22	uptake.ab,ti.
23	utility.ab,ti.
24	utilization.ab,ti.
25	utilisation.ab,ti.
26	credibility.ab,ti.
27	fit.ab,ti.
28	relevance.ab,ti.
29	"compatib*".ab,ti.
30	"suitab*".ab,ti.
31	usefulness.ab,ti.
32	reach.ab,ti.
33	spread.ab,ti.
34	coverage.ab,ti.
35	continuation.ab,ti.
36	"durab*".ab,ti.
37	"incorporat*".ab,ti.
38	"integrat*".ab,ti.
39	institutionalisation.ab,ti.
40	institutionalization.ab,ti.

14

41	routinization.ab,ti.	
42	routinisation.ab,ti.	
43	satisfaction.ab,ti.	
44	agreeable.ab,ti.	
45	discontinuation.ab,ti.	
46	de-adoption.ab,ti.	
47	normalisation.ab,ti.	
48	normalization.ab,ti.	
49	(implement* adj3 cost).ab,ti.	
50	"internal consistency".ab,ti.	
51	test-retest.ab,ti.	
52	"test retest".ab,ti.	
53	(reliability and (interrater or inter-rater or intrarater or intra-rater)).ab,ti.	
54	"content validity".ab,ti.	
55	"face validity".ab,ti.	
56	"construct validity".ab,ti.	
57	"criterion validity".ab,ti.	
58	"structural validity".ab,ti.	
59	"concurrent validity".ab,ti.	
60	"predictive validity".ab,ti.	
61	"convergent validity".ab,ti.	
62	"discriminant validity".ab,ti.	
63	"principal components analys*".ab,ti.	
64	"factor analys* ".ab,ti.	
65	"factor structure* ".ab,ti.	
66	dimensionality.ab,ti.	
67	"Item response model".ab,ti.	
68	"Item response theory".ab,ti.	
69	IRT.ab,ti.	
70	MIMIC.ab,ti.	
71	"classical test theory".ab,ti.	
72	EFA.ab,ti.	
73	CFA.ab,ti.	
74	(exploratory or confirmatory).ab,ti.	
75	factor.ab,ti.	
76	74 and 75	
77	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10	
78	11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49	

15

79	50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 76
80	77 and 78 and 79
81	exp animals/ not humans.sh.
82	80 not 81

16

Table 2. Implementation outcomes and their synonyms

Acceptability	acceptab*, agreeable, satisfaction, credibility
Adoption	adopt*, uptake, utility, utilization, utilisation, discontinuation, de-adoption
Appropriateness	appropriate*, fit, relevance, compatib*, usefulness
Feasibility	feasib*, suitab*, practicability, applicab*, workab*, transferab*
Implementation cost	Cost
Penetration	penetrat*, reach, spread, coverage
Sustainability	sustain*, maintenance, continuation, durab*, incorporat*, integrat*, institutionalisation, institutionalization, routinization, routinisation, normalisation, normalization

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

NS is the Director of London Safety and Training Solutions Ltd, which provides quality and safety training and advisory services on a consultancy basis to healthcare organization globally. The other authors declare that they have no competing interests.

Authors' contributions

Zarnie Khadjesari (ZK) designed and drafted the protocol, and is guarantor of the review. Zarnie Khadjesari, Louise Hull (LH) and Nick Sevdalis (NS) conceived the study. ZK, LH and Silia Vitoratou (SV) piloted the data extraction forms. SV is developing the instrument quality criteria. All authors provided feedback on the review methods and contributed to the final manuscript.

Acknowledgements

18

19

REFERENCES

1. Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. Implement Sci. 2009;4:18.

2. Wilson PM, Sales A, Wensing M, Aarons GA, Flottorp S, Glidewell L, et al. Enhancing the reporting of implementation research. Implement Sci. 2017;12:13.

3. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation Hybrid Designs. Med Care. 2012;50(3):217–26.

4. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci. 2013;8:139.

5. Grimshaw J, Eccles M, Thomas R, MacLennan G, Ramsay C, Fraser C, et al. Toward evidence-based quality improvement. Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966-1998. J Gen Intern Med. 2006;21 Suppl 2:S14-20.

6. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation Research in Mental Health Services: an Emerging Science with Conceptual, Methodological, and Training challenges. Adm Policy Ment Health. 2009;36(1).

7. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011;38(2):65–76.

8. Powell BJ, Beidas RS, Lewis CC, Aarons GA, McMillen JC, Proctor EK, et al. Methods to Improve the Selection and Tailoring of Implementation Strategies. J Behav Health Serv Res. 2017;44(2):177–94.

9. Emmons KM, Weiner B, Fernandez ME, Tu S-P. Systems antecedents for dissemination and implementation: a review and analysis of measures. Health Educ Behav. 2012;39(1):87–105.

10. Scott T, Mannion R, Davies H, Marshall M. The Quantitative Measurement of Organizational Culture in Health Care: A Review of the Available Instruments. Health Serv Res. 2003;38(3):923–45.

11. King T, Byers JF. A review of organizational culture instruments for nurse executives. J Nurs Adm. 2007;37(1):21–31.

12. Weiner BJ, Amick H, Lee S-YD. Review: Conceptualization and Measurement of Organizational Readiness for Change: A Review of the Literature in Health Services Research and Other Fields. Med Care Res Rev. 2008;65(4):379–436.

13. Squires JE, Estabrooks CA, Gustavsson P, Wallin L. Individual determinants of research utilization by nurses: a systematic review update. Implement Sci. 2011;6:1.

14. Chor KHB, Wisdom JP, Olin S-CS, Hoagwood KE, Horwitz SM. Measures for Predictors of Innovation Adoption. Adm Policy Ment Health Ment Health Serv Res. 2015;42(5):545–73.

20

15. Chaudoir SR, Dugan AG, Barr CHI. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. Implement Sci. 2013;8:22.

16. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci. 2009;4:50.

17. Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. Implement Sci. 2015;10:155.

18. Clinton-McHarg T, Yoong SL, Tzelepis F, Regan T, Fielding A, Skelton E, et al. Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: a systematic review. Implement Sci. 2016;11:148.

19. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci. 2015;10.

20. Terwee CB, Prinsen C a. C, Ricci Garotti MG, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res. 2016;25(4):767–79.

21. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651–7.

22. Glasgow RE, Riley WT. Pragmatic measures: what they are and why we need them. Am J Prev Med. 2013;45(2):237–43.

23. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;350:g7647.

24. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.

25. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res. 2009;18(3):313–33.

26. Lewis CC, Stanick CF, Martinez RG, Weiner BJ, Kim M, Barwick M, et al. The Society for Implementation Research Collaboration Instrument Review Project: A methodology to promote rigorous evaluation. Implement Sci. 2015;10:2.

27. The Health Foundation. Evidence scan: Improvement Science. London: The Health Foundation; 2011.

28. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2008.

29. Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. J R Stat Soc Ser A. 2002;165(2):233–53.

21

30. Rogers E M. Diffusion of innovations. 4th edition. New York: The Free Press; 1995.

31. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.

32. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. Implement Sci. 2014;9:118.

33. Aarons GA, Fettes DL, Sommerfeld DH, Palinkas L. Mixed Methods for Implementation Research: Application to Evidence-Based Practice Implementation and Staff Turnover in Community Based Organizations Providing Child Welfare Services. Child Maltreat. 2012;17(1).

34. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115–23.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Manuscript page number
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	N/A
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	18
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	6
Support:			
Sources	5a	Indicate sources of financial or other support for the review	18
Sponsor	5b	Provide name for the review funder and/or sponsor	N/A
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	18
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-5
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	6
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	7-9
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	7

Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	6-7, 14-16
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	10
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	10
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	10-11
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	10-11
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	11
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	9
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	N/A
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	N/A
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	N/A
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	11
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	N/A
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	N/A

* It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.