

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

Zarnie Khadjesari,¹ Silia Vitoratou,² Nick Sevdalis,¹ Louise Hull¹

To cite: Khadjesari Z, Vitoratou S, Sevdalis N, *et al.* Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol. *BMJ Open* 2017;7:e017972. doi:10.1136/bmjopen-2017-017972

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-017972>).

Received 29 May 2017
Revised 8 September 2017
Accepted 14 September 2017



CrossMark

¹Department of Health Service and Population Research, Centre for Implementation Science, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK

²Department of Biostatistics and Health Informatics, Psychometrics and Measurement Lab, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK

Correspondence to

Dr Zarnie Khadjesari;
zarnie.khadjesari@kcl.ac.uk

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 measures 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 and (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 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.

Trial registration number International Prospective Register of Systematic Reviews (PROSPERO): CRD42017065348.

Strengths and limitations of this study

- We have designed a comprehensive search strategy for published and unpublished literature and have included 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.
- Due to the breadth of the setting (ie, 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.

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 (ie, 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; and (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 focused on measures of system-level antecedents to implementation,⁹ organisation-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, a meta-theoretical framework that aims to understand and/or explain influences on implementation outcomes.^{16–18} Furthermore, a review has focused 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 (ie, excluding instruments specific to mental healthcare settings). The working taxonomy of implementation outcomes is relevant across stakeholder levels and stages of implementation, and can be applied to different implementation models, theories and frameworks.¹⁹ This review will complement and allow direct comparison with the review by Lewis *et al*, which 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, that is, 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 1-year period and found that only 59% had searched EMBASE (where searching MEDLINE and EMBASE databases is considered a minimal requirement by the authors²⁰), 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 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.

Table 1 Search strings for MEDLINE

SI No	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.

Continued

Table 1 Continued

SI No	Search strings for MEDLINE
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 intrarater)).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
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

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 (ie, implementation literature), (2) the constructs being measured (eg, adoption) and (3) the measurement properties of instruments (eg, 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 science²⁷ and index terms (eg, MeSH) applied to Lewis *et al*'s published systematic review protocol²⁶ 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 that they included all relevant synonyms. We will also conduct a supplementary search for the names of the instruments that 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 and 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 Hand²⁹). Psychometric scales consist of items that 'do not alter or influence the underlying concept: they are merely aspects of it, or indicators of its magnitude' (Fayers and Hand, p236)²⁹, whereas clinimetric scales consist of items that are 'merely constructing an index [...] and need not to be indicator variables for the concept in question' (Fayers and Hand, p237).²⁹ These instruments may consist of surveys, checklists and/or questionnaires, 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 or 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,

Table 2 Implementation outcomes and their synonyms

Implementation outcomes	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

appropriateness, adoption, penetration, fidelity, implementation cost and sustainability.⁷ For each outcome, they suggest the level of analysis (eg, organisation, provider, consumer), theoretical basis (eg, Rogers' theory of the diffusion of innovation³⁰), overlapping constructs, salient implementation stage (eg, early for adoption, ongoing for penetration, late for sustainability) and suitable research methods for measurement (eg, 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 (eg, preimplementation, throughout implementation, postimplementation). Implementation outcomes may focus on attitudes, knowledge, behaviours, costs or number of participants receiving an intervention, among others.

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 (ie, not focused on a specific practice,¹⁷ such as contingency management). These criteria were needed 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 the 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 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 and 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 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

Predesigned 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, level of analysis (ie, organisation, provider, consumer), focus of measure (eg, attitudes, knowledge, behaviour or other) and implementation stage (eg, preimplementation, throughout implementation, postimplementation).

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, name of instrument and version, type of psychometric study, setting, sample characteristics (eg, gender, age, ethnicity), characteristics of the intervention or innovation being implemented, 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 infeasible (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 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 and 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 semistructured 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. While 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.

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. 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.

Acknowledgements We are enormously grateful to Dr Cara Lewis, Professor Bryan Weiner, Caitlin Dorsey and Kayne Mettert for their time and generosity in sharing their valuable review experience with us. We have learnt a huge amount from them and believe our similar approach will benefit the field by enabling comparison between the findings. We would also like to thank information specialist Kate Lewis-Light for her advice on the search strategy and our stakeholder groups for their helpful feedback on the search terms, in particular: Professor Annette Boaz (Professor of Health Care Research), Dr Lucy Goulding (Post-Doctoral Service Improvement Specialist, King's Improvement Science), Dr Heidi Lempp (Senior Lecturer in Medical Sociology), Professor Brian Mittman (Senior Advisor, VA Center for Implementation Practice and Research Support; Consultant, University of California, Los Angeles Institute for Innovation in Health; Senior Advisor, RAND Health), Professor Andrew Pickles (Chair in Biostatistics), Professor Nigel Pitts (Director of Innovation and Implementation), Professor Jane Sandall (Chair in Social Science and Women's Health and Collaboration for Leadership in Applied Health Research and Care South London theme lead for maternity and women's health and capacity building) and Dr Bryony Soper (Honorary Professor, Brunel University).

Contributors ZK designed and drafted the protocol and is guarantor of the review. ZK, LH and NS conceived the study. ZK, LH and 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.

Funding NS' research is funded by the National Institute for Health Research (NIHR) via the 'Collaboration for Leadership in Applied Health Research and Care South London' (CLAHRC South London) at King's College Hospital National Health Service (NHS) Foundation Trust, London, UK. ZK and LH are funded by King's Improvement Science, which is part of the NIHR CLAHRC South London and comprises a specialist team of improvement scientists and senior researchers based at King's College London. King's Improvement Science is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College

Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), Guy's and St Thomas' Charity, the Maudsley Charity and the Health Foundation. NS is also funded by the South London and Maudsley NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of NHS, NIHR or the Department of Health.

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 organizations globally. The other authors declare that they have no competing interests.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Eccles MP, Armstrong D, Baker R, *et al*. An implementation research agenda. *Implement Sci* 2009;4:18.
2. Wilson PM, Sales A, Wensing M, *et al*. Enhancing the reporting of implementation research. *Implement Sci* 2017;12:13.
3. Curran GM, Bauer M, Mittman B, *et al*. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care* 2012;50:217–26.
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, *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, *et al*. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. *Adm Policy Ment Health* 2009;36:24–34.
7. Proctor E, Silmere H, Raghavan R, *et al*. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health* 2011;38:65–76.
8. Powell BJ, Beidas RS, Lewis CC, *et al*. Methods to Improve the Selection and Tailoring of Implementation Strategies. *J Behav Health Serv Res* 2017;44:177–94.
9. Emmons KM, Weiner B, Fernandez ME, *et al*. Systems antecedents for dissemination and implementation: a review and analysis of measures. *Health Educ Behav* 2012;39:87–105.
10. Scott T, Mannion R, Davies H, *et al*. The quantitative measurement of organizational culture in health care: a review of the available instruments. *Health Serv Res* 2003;38:923–45.
11. King T, Byers JF. A review of organizational culture instruments for nurse executives. *J Nurs Adm* 2007;37:21–31.
12. Weiner BJ, Amick H, Lee SY. 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;65:379–436.
13. Squires JE, Estabrooks CA, Gustavsson P, *et al*. Individual determinants of research utilization by nurses: a systematic review update. *Implement Sci* 2011;6:1.
14. Chor KH, Wisdom JP, Olin SC, *et al*. Measures for predictors of innovation adoption. *Adm Policy Ment Health* 2015;42:545–73.
15. Chaudoir SR, Dugan AG, Barr CH. 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, *et al*. 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, *et al*. 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, *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;10:53.
20. Terwee CB, Prinsen CA, Ricci Garotti MG, *et al.* The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res* 2016;25:767–79.
21. Terwee CB, Mokkink LB, Knol DL, *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
22. Glasgow RE, Riley WT. Pragmatic measures: what they are and why we need them. *Am J Prev Med* 2013;45:237–43.
23. Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
24. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
25. Mokkink LB, Terwee CB, Stratford PW, *et al.* Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res* 2009;18:313–33.
26. Lewis CC, Stanick CF, Martinez RG, *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 CRD. *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 Stat Soc* 2002;165:233–53.
30. Rogers EM. *Diffusion of innovations*. 4th edition. New York: The Free Press, 1995.
31. Liberati A, Altman DG, Tetzlaff J, *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, *et al.* 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.
34. Terwee CB, Jansma EP, Riphagen II, *et al.* Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.