Developing Standards for Reporting Phase IV Implementation studies

(StaRI)

Investigators

Hilary Pinnock[1]
Steph Taylor[2]
Eleni Epiphaniou[2]
Aziz Sheikh[1]
Chris Griffiths[2]
Sandra Eldridge[2]
Steff Lewis[1]
Peter Craig[3]

[1] Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, UK

[2] Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, London, UK

[3] Chief Scientist Office, Scottish Government Health & Social Care Directorates, St Andrew's House, Regent Road, Edinburgh, EH1 3DG

Table of contents

1. Background

1.1 Phase IV implementation studies

The Medical Research Council (MRC) provides a framework to help researchers, funders and other decision makers make appropriate methodological decisions for designing, evaluating and implementing complex interventions.[1] Phase III trials, i.e. Randomised Controlled Trials (RCTs), are the gold standard of research designs for assessing the efficacy/effectiveness of interventions.[2] RCTs are delivered under tightly controlled conditions, with carefully selected, highly motivated, fully informed and consented participants, and follow rigid protocols to avoid the influence of confounding variables and limit the impact of bias,[3,4] though pragmatic trials and cluster randomised trials may avoid or mitigate some of these issues. Phase IV studies which accommodate – or even encourage - the diversity of patient, professional and healthcare contexts in order to inform implementation in real-life settings are relatively uncommon.[5] As well as assessing their effectiveness at the population level,[6] health care professionals need practical information about the impact on time and resources, the training requirements and workplace implications of implementing interventions into routine care.[4,7]

Phase IV studies aim to evaluate the implementation by healthcare services of research findings and, more specifically, to translate interventions that have proven to be clinically- (and increasingly cost-) effective in research settings into routine care. They compare the new procedure with the existing/previous regime and explore whether it improves patient outcomes, quality of care, service delivery and/or the health and social well-being of the population[8] and determine the intervention's true population effect.[9] We will use the definition described by the *Journal of Implementation Science* which defines implementation research as:

> 'The scientific study of methods to promote the systematic uptake of proven clinical treatments, practices, organisational, and management interventions into routine practice, and hence to improve health. In this context, it includes the study of influences on patient, healthcare professional, and organisational behaviour in either healthcare or population settings'.[10]

Effectiveness of an intervention is assessed in a heterogeneous, unselected population and real-life clinical settings, and examines outcomes relevant to the patient, provider, social and healthcare contexts.[4,11] The focus is on external validity and generalisability where the effectiveness to the population's social and health well-being is assessed. Thus, they are useful study designs when developing policy recommendations.[8]

1.2 The (poor) quality of reporting of implementation studies

The importance of developing standards for transparent and accurate reporting of implementation studies has been recently highlighted.[12,13] In addition, as part of the Practical systematic Review of Self-Management Support for long-term conditions (PRISMS) study,[5] we recently conducted a systematic review of the evidence on routine implementation of self-management support interventions in real-life populations with one or more exemplar long-term conditions. We specifically wanted to identify Phase IV implementation studies in which there was a comparator

group (including randomised/non-randomised allocation of groups/services, 'before and after', interrupted time series, step wedge designs).

At different stages of the review we had significant challenges related to:

- The inconsistent and diverse terminology. Echoing one study which identified 29 terms in current used to describe studies which aimed to translate research evidence into action,[14] we found that the plethora of descriptors used for implementation studies made it difficult to identify terms for our search strategy. After scrutinising the key words and titles of a number of implementation-related papers, we included a range of terms including 'effectiveness trials', 'routine clinical care', 'implement*', 'real-world', 'Phase IV', 'pragmatic'. This however, proved to be non-specific as these terms were used (seemingly arbitrarily) for a wide range of study types. For example, Kemple et al.,[15] described their study as a 'single-blinded RCT' with no mention of the word 'implementation', though the intervention was at a service level with no individual patient recruitment. In contrast, Carter et al.,[16] used the term 'implementation' to describe their RCT of a self-management diabetes intervention in which patients were individually randomised. Sämann et al., however, used the word 'implementation' appropriately in their series of studies to describe their evaluation of a diabetes teaching and training programme implemented in routine practice.[17-19] Finally, Homer et al.,[20] used the terminology 'quality improvement' rather than 'implementation' and the abstract implies that patients were randomised though the full-text is clear that randomisation was at service level.

- Lack of clarity about whether recruitment was to the service (as would be expected in an implementation study) or to the study (as is the norm in a RCT), often made it difficult to determine whether the study was Phase IV even when reading the full text. For example, Gruesser et al.,[21] are clearly evaluating the implementation of an educational intervention for people with hypertension, but even when reading the full text it is not clear how patients were recruited to the programme and whether recruitment was conditional on participating in research. Delaronde et al.,[22] used a preference design within a managed care organisation, so that clinically eligible patients could opt in or opt out. Those not expressing a clear preference were randomised, though it is not clear how (or whether) the research aspect of the evaluation were explained.

- Poor reporting of implementation studies. Although study aims and objectives, description of the main outcomes and the clinical intervention were often well – or at least adequately – described, there were many areas where implementation information was missing or was unclear. Examples included lack of details on training for professionals delivering the new service,[23-25] unclear descriptions (numbers and characteristics) of the whole population,[26-28] and, within this population, of those eligible for the service,[26,27] no/inadequate reporting of data service uptake, numbers of withdrawals and possible reasons for withdrawal.[19,25] Table 1 below provides a brief description of the number of studies failing to report the above information in the PRISMS implementation systematic review.

Table 1: Frequency of poor reporting in papers about Phase IV self-management support interventions included in the PRISMS systematic review.[5]

Reporting standards	% of studies **not** reporting this	Notes
Uptake of the intervention is a key outcome	60/62 = 96.8%	Uptake was often calculated by the reviewers for studies reporting whole population and participating population
Some, and ideally the primary outcome, should be at population level (normally using routinely collected data)	54/61 = 88.5%	63.9% reported routinely collected data but not all were at eligible population level
There must be a full description of the setting	42/61 = 68.8%	Settings involved the hospital, the community or the clinics
Report on eligible population characteristics	41/61 = 67.2%	-
Report of training for professionals	38/61 = 62.3%	-
Reasons for withdrawal	36/61 = 59%	-
Report of whole population numbers	29/61 = 47.5%	-
Number of withdrawals	28/61 = 45.9%	-
Based explicitly on Phase III evidence and/or guideline recommendations	24/61 = 39%	-
Report of eligible population numbers	23/61 = 37.7%	-
Eligibility is for the service, NOT the research	19/61 = 31.15%	Despite the majority reporting eligibility criteria it was unclear for many studies whether participants were recruited to the service/programme or study
Practical information about how the intervention was implemented, as well as a description of the intervention	9/61 = 14.7%	-

1.2.1 Lessons learned from implementation studies included in the PRISMS review

- Routinely collected data, at a population level, is an important element to be reported when conducting implementation studies. However, some authors questioned the reliability of routinely collected data due to data-entry inconsistencies (incomplete, underestimated or overestimated data reporting),[22,24] highlighting the importance of improving procedures for standardised recording of routine clinical data.[29]

- Implementing complex interventions in routine care invariably requires (sometimes substantial) changes to the organisation/service. The service context and any

reconfiguration need to be reported and described in detail, both to enable readers to assess applicability to their own practice, but also to assess the process by which an intervention has an impact.[30,31]

- Implementing a complex intervention into diverse settings means that aspects of the intervention will be adapted to suit individual patients, providers, and service cultures, and may (indeed probably should) evolve over time.[32] Fidelity to core aspects of the intervention, however, should be checked and this process of 'normalisation' (or not) described.[33]

- Journal editors are increasingly expecting authors to complete relevant reporting standard checklists when submitting papers. The absence of a checklist for implementation studies may well lead some authors to 'shoe-horn' their study into the CONSORT requirements for reporting an RCT. This not only results in inappropriate requirements (e.g., individual eligibility and recruitment to a trial) but also leads omission of important criteria for reporting implementation studies (e.g., uptake of the intervention, outcomes from routine clinical data). This is also potentially important for funding bodies and grant reviewers who are considering applications for Phase IV studies.

- The expectation is that Phase IV studies would build on Phase III evidence (and/or guideline recommendations based on Phase III evidence) and that they would reflect on their findings in this light, specifically compare and comment on effect size and any planned or unplanned effects compared to the original evaluation.[1] However, only a few studies explicitly did this (e.g implementing a telephone asthma review service halved the effect of the Phase III RCT[34])

1.3 Existing EQUATOR reporting standards

Introduction of (for example) the CONSORT, COREQ and PRISMA checklists have standardised the reporting of RCTs, qualitative studies and systematic reviews.[35-37] Similarly, guidelines have been developed for reporting observational studies[38] quality improvement studies,[39] and non-randomised public health interventions.[40] Our experience in the PRISMS systematic review has emphasised the need similarly to improve the reporting of Phase IV implementation studies.

Table 2 below provides a brief summary of differences, similarities and overlaps between StaRI and other checklists.

Table 2: Other checklists and their differences/similarities with StaRI

Guideline	Overlap with StaRI	Useful items for StaRI	Differences with StaRI
STROBE – for reporting observational studies and more specifically cohort, case-control and cross-sectional studies.[38]	Minimal: STROBE is concerned with comparing 'exposed' with 'not-exposed' at an individual patient level. StaRI implements in a population/service and uptake is an outcome.	Some generic reporting standards will be relevant (e.g., describing context, describing eligibility, defining outcomes)	StaRI will need to extend these generic standards to describe impact on the service and reconfiguration, turnover in the eligible population over time, uptake of the new service.

Guideline	Overlap with StaRI	Useful items for StaRI	Differences with StaRI
TREND - refers to non-randomised trials relevant to behavioral and public health interventions.[40]	Some overlap as non-randomised designs may be used for implementation studies. However, TREND may be evaluating previously untested Phase III interventions (typically at public health level)	Many of the descriptive standards (Items 3 and 4) may be useful and aspects of reducing bias (item 8) may be relevant. Comparison of population at baseline and follow up is relevant.	The Phase III experimental basis of the intervention is important to StaRI. The emphasis on 'non-compliance' is not appropriate (though 'uptake' is a key outcome). The discussion should consider the impact on the service configuration.
CONSORT extension for reporting pragmatic trials - effectiveness of trial interventions in routine care.[41]	Some overlap as the CONSORT extension is also concerned with routine clinical care, but these may be at Phase III level whereas StaRI is at Phase IV.	Many of the descriptive items will be useful (Item 3 and 4) especially the reporting of diversity in how an intervention is delivered in different healthcare settings.	At phase IV participant eligibility is to the service not trial. Randomisation (if relevant) is at service level, not patients). Key outcomes are likely to involve routine clinical data and include service uptake
SQUIRE - a checklist developed to improve the completeness, transparency and accuracy of the reporting of processes aiming to address local problems or improve 'dysfunctional' systems.[39]	Some overlap as operating at service level, but the local cycle of auditing / planning / implementing / re-evaluating / reflexivity is not a necessary component of implementation studies.	Many of the descriptive items will be useful especially the description of how the intervention was embedded in the service.	The pre-requisite for Phase IV studies is Phase III evidence rather than a local problem/dysfunctional service. Use of routine clinical data is not highlighted in SQUIRE but is crucial to StaRI

Thus, whilst STROBE, SQUIRE, TREND and the extension of CONSORT for Pragmatic Trials have some relevance, they are not wholly applicable to Phase IV implementation studies. We thus aim to develop a checklist for quality standards when reporting a Phase IV implementation study which researchers can use to improve completeness and transparency of their reporting, and editors can apply to assess the quality of publications. In additional benefit reporting standards will raise the profile of implementation research, potentially stimulate interest in the methodology and make it easier to identify relevant papers to inform development of healthcare services.

1.3.1 Guidelines under development

- CReDECI: To ensure high quality reporting of the development and evaluation of complex intervention a set of criteria based on the recent update of the MRC framework has been developed. However, this criteria list covers only the first three stages of the MRC framework i.e. (1) development, (2) feasibility and piloting, (3) intervention and evaluation.[42] The final Phase IV of the MRC framework will not covered by this checklist.

- RECORD: The REporting of studies Conducted using Observational Routinely-collected Data statement.[43] This is an extension to the STROBE statement to ensure transparent methods of reporting routine clinical data. It is likely that routinely collected data will be important in reporting whole population data for Phase IV implementation study and some of the standards of the RECORD checklist may be useful

2. Aims and objectives

2.1.1 Aim

We aim to develop a guideline to improve the quality of reporting Phase IV implementation studies, specifically those with a comparator group.

2.1.2 Objectives

1. To review the literature related to designing and reporting Phase IV implementation studies in order to identify existing standards
2. To recruit an international expert panel and conduct a Delphi exercise to identify and prioritise standards for reporting Phase IV implementation studies
3. To convene a consensus meeting to agree the items for the checklist and content of a guideline for reporting Phase IV implementation studies

3. Methods

We will follow the methodology described in the guidelines for Developing Health Research Reporting Guidelines.[44]

3.1 Review the literature related to designing and reporting Phase IV implementation studies in order to identify existing standards

3.1.1 Existing guidance

In the context of reporting Phase IV implementation studies, we will undertake a literature review to identify:

- evaluations of current design and reporting practice
- existing advice, guidance, frameworks, standards

We will search the MEDLINE database, using the guideline terms such as *'standard*'*, *'guidance'; 'framework'; 'reporting guideline*'; (report* ADJ GOOD ADJ PRACTICE)* AND study design such as *'implementation'; 'implementation science'; 'Phase IV'; 'Phase 4'; 'real-life'; 'routine clinical care'; 'Real-world or 'real world' or routine or nationwide) adj1 (setting* or practice or context)*. We will explore existing statement from existing EQUATOR guidance and undertake snowball searches from their reference lists, and hand search the *Journal of Implementation Science*, *Pragmatic and Observational Research*, *Quality and Safety in Healthcare.*

Identified studies will be scrutinised for potential standards for inclusion in the StaRI statement

3.1.2 Quality of reporting in publications of Phase IV implementation studies

We have recently undertaken a systematic review of implementation studies in the context of asthma self-management support (the PRISMS study) which highlighted many deficiencies.[5] For example: inconsistent terminology, no explicit under-pinning phase III work or guideline recommendations, no reporting of whole (eligible) population data, unclear whether eligibility/recruitment is for the service (as opposed to the research), no outcomes reported at population level (normally using routinely collected data), uptake of /retention in the intervention not reported, representativeness of sub-groups in relation to the whole population not described, poor description of the setting and how the intervention was implemented, fidelity to crucial components of the intervention not assessed.

We will systematically identify these deficiencies from our PRISMS systematic review, and other systematic reviews of implementation studies identified in our literature review and collate them as possible standards for inclusion in a StaRI statement.

3.2 Recruit an international expert panel and conduct a Delphi exercise to identify and prioritise standards for reporting Phase IV implementation studies

3.2.1 Recruit an international expert panel

We will recruit an international expert panel (approx. n=30) to include representatives of:

- International methodologists identified in the course of our literature review, specifically including professionals involved with the MRC framework for the design and evaluation of complex intervention [1] and EQUATOR.[35]
- Journal editors from high impact general (e.g., *BMJ*, *PLoS Med*, *Lancet*) and methodology-specific journals (e.g., *J Implementation Science, Pragmatic Observational Res*)
- Funding bodies (National Institute for Health Research, Economic and Social Research Council, Medical Research Council, Wellcome Trust; Federal Ministry of Education and Research, (Germany); National Institutes of Health, (USA); National Health and Medical Research Council, (Australia); Health Research Council (New Zealand); European Union, WHO; European Union 7th Framework program. CORDIS (Community Research and Development Information Service); Chief Scientist Office (Scotland).
- We will also include charities who fund research (Asthma UK, Cancer Research UK, Diabetes UK, Astma Fonds)
- International researchers who have published high profile implementation research

We will e-mail potential members of the expert panel, inviting them to participate. The invitation will include a description of the process, the anticipated timescale, and the estimated commitment. We will invite participants to commit to either participation in the e-Delphi exercise, and/or the workshop.

3.2.2 Conduct a Delphi Exercise

Originating from the RAND Corporation in the 1950s,[45] the Delphi method is a technique for reaching consensus amongst an expert panel.[46-51] The underlying concept is that an expert panel is recruited who contribute ideas, and then rank suggestions in successive rounds until pre-defined consensus is reached. The panellists work independently and their contributions are anonymous, but in each round responses are influenced by summary feedback from previous rounds. As face-to-face discussion is not required, the exercise can be administered by e-mail. The technique has been widely used in a range of health care contexts including defining the components of an anaphylaxis plan,[52] identifying safety standards of GP computer systems,[53] prioritising research needs within the UK,[54] and internationally.[55]

Crucial to the methodology is the commitment of the participants to complete all rounds of the exercise so that the outcomes reflect the group consensus. This will be made clear in the invitation and the process will be streamlined to reduce the burden on participants and timelines designed to be sufficiently generous to allow busy colleagues to respond whilst being sufficiently tight to maintain interest. We aim to recruit at least 30 participants to the consensus exercise which will be conducted by e-mail. We will use Clinvivo systems to facilitate the e-Delphi process. [www.clinvivo.com]

Open round

The first step is to ask open an open question and invite the expert panel to contribute ideas for ranking in the subsequent rounds,[46,49] We will ask participants to suggest at least six standards which should be required in reporting Phase IV implementation trials. To aid deliberation and collation of the responses we will suggest that responses should be considered under appropriate headings (e.g., rationale and underpinning evidence for the study, description of setting, recruitment, intervention, outcomes and data collection, presentation of results, and interpretation), with an additional space for providing suggestions which do not fit into, or fit into more than one category. The question, and categories for the responses, will be refined and piloted to ensure optimal terminology and clarity.

Standards derived from the literature review will be provided as examples, but it will be emphasised that these are only provided to assist clarity in the task and to prompt thinking. Participants will be asked to indicate if any of the suggested standards are particularly important to them.

Responses will be collated and discussed by the research team and a checklist of possible standards derived for scoring in Round 1.

Scoring round 1

Panel members will be asked individually to score each item on the checklist on a score of 1 to 9. (least important to very important). There will be an opportunity at the end of the checklist to add any further standards which respondents feel should be considered.

Participants will be asked to return their completed questionnaires within 2 weeks, with a reminder being sent a few days before the deadline. For participants not returning their questionnaires 2

weeks post-deadline a reminder email will be sent the day after the due day with a 7-day deadline. Results will be collated and the median score and percentage of agreement calculated.

Scoring round 2

Participants who completed round 1 will be sent a round 2 checklist in which the median results from the first scoring round will be listed alongside the participant's own score. To aid clarity we will consider grouping standards according to their ranking in Round 1 (e.g., '80% agreement with scores 7, 8 or 9' or '80% agreement with scores 1, 2 or 3'). Participants will be invited to reconsider the importance of the standards and confirm or revise their score in the light of the group opinions.

Participants will be asked to return their completed questionnaires within 2 weeks, with a reminder being sent a few days before the deadline. For participants not returning their questionnaires 2 weeks post-deadline a reminder email will be sent the day after the due day with a 7-day deadline. Results will be collated using Excel to calculate the median score and percentage of agreement.

Reaching agreement

Consensus is defined as 80% agreement for the priority score of 7, 8 or 9. We anticipate that two scoring rounds will allow an acceptable degree of agreement on research priorities, but if not a final third scoring round will be undertaken. This will follow the format of round 2, but omit items which had 80% agreement with scores 1, 2 or 3.

3.3 Convene a consensus workshop to agree the items for the checklist and content of a guideline for reporting Phase IV implementation studies

We will seek funding for an international consensus workshop at which we will discuss individual items and agree the first draft of the StaRI checklist. The final list of items for discussion at the consensus workshop will be decided by the study team based largely on the outcome of the Delphi exercise, but also taking other published standards for implementation studies into account.[44]

3.3.1 Pre-meeting materials

Pre-meeting material will be sent to delegates which will include an overview of the background information, the results of the Delphi exercise and the list of candidate items for the checklist.

3.3.2 The consensus meeting

We will select a convenient venue (ideally in connection with an international meeting to facilitate travel) and convene a 2 day workshop. Consideration will be given to arranging videoconferencing facilities to enable international delegates.

Participants

We will invite up to 20 participants to contribute to the workshop including methodologists, experienced health service researchers, journal editors and representatives of funding bodies.

Agenda

After the initial presentations the delegates will discuss the candidate items and reach a decision on those that should be included (see table 3). The workshop discussions will be audio recorded with consent.

- Each item on the checklist will be discussed in the light of the e-Delphi prioritisation and any relevant evidence, and discussed until consensus is reached on whether the standard should be included (or not). In the event that agreement cannot be reached a vote may be taken.[44]
- Discuss whether there is a need for a template flow diagram, and if so consider its contents and create an initial draft
- Discuss and plan the publications (StaRI statement, explanatory document), agree key tasks and timelines.

Table 3. Tasks and outputs of the workshop

The workshop will begin with the presentation of the background literature, the results of the Delphi process and an overview of the tasks		
Task	Process	Output
Draft checklist	Discuss the rationale for including each of the items using a combination personal experience and (where available) empirical evidence to support (or not) the importance of incorporating the item into the checklist	A first draft of the checklist for reporting implementation studies
Draft flow diagram	Discuss the development and content of a flow diagram for reporting implementation studies.	Draft flow diagram
Plans	Agree the on-going strategy for producing the checklist, publication and an explanatory paper,	Agreed strategy
Dissemination strategy	Agree publication plans, and other plans for disseminating the guideline	

3.4 Production of the StaRI statement and dissemination

3.4.1 Production of the reporting guideline statement

Based on the consensus workshop, the team of investigators (with co-opted expertise as appropriate) will take the items selected by delegates and draft the checklist and statement to be finalised in e-mail discussion with members of the consensus group.[44]

3.4.2 Dissemination

A publication strategy and a proactive approach to dissemination will be agreed at the consensus workshop but will include a statement published in a high profile peer review journal, with a checklist and (probably) a flow chart. In addition to the StaRI statement for publication, consideration will be given to writing a comprehensive 'Explanation and Elaboration' document to explain the rationale for the items in the checklist.[44]

3.5 Organisation

Dr Hilary Pinnock and Professor Steph Taylor (co-PIs) will supervise the project. Dr Eleni Epiphaniou will undertake the literature search, setting up and piloting of the Delphi exercise (with administrative support for running the three rounds of the Delphi) and organisation of the consensus workshop. The team of investigators, with co-opted support as required from the Expert Panel, will support the development of the Delphi questionnaire, the interpretation of data, preliminary work for the consensus workshop and preparation of the final statement and supporting documentation.

3.5.1 Timeline

Months	Tasks
July-September 2013	Register protocol with EQUATOR Undertake literature review Recruit members of the expert panel Pilot the Delphi exercise
October 2013 – March 2014	Delphi exercise: Open round, Scoring round 1, Scoring round 2 (+ third round if necessary
May 2014	Consensus workshop
June – September 2014	Draft, circulate and finalise publications Dissemination

3.5.2 Funding

The work for objectives 1 and 2 have been funded in part by internal funding from QMUL. HP is supported by a Primary Care Research Career Award from the Chief Scientist's Office of the Scottish Government and this grant contributed to the research costs. Funds will be needed to support the workshop.

3.6 References

1. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. London: MRC; 2008. Available from www.mrc.ac.uk/complexinterventionsguidance [accessed August 2013]
2. Garfield SA, Malozowski S, Chin MH, Narayan KM, Glasgow RE, Green LW, Hiss RG, Krumholz HM for the DMICC. Considerations for diabetes translational research in real-world settings. *Diabetes Care* 2003;26:2670-4

3. Bhattacharyya O, Reeves S, Zwarenstein M. What Is Implementation Research? Rationale, Concepts, and Practices. *Res Social Work Prac* 2009;19:491-502

4. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to- effectiveness transition. *Am J Public Health* 2003;93:1261-7

5. Taylor S, Pinnock H, Sheikh A, Griffiths C, Greenhalgh T, Ribchester J. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions. http://www.netscc.ac.uk/hsdr/projdetails.php?ref=11-1014-04 [accessed August 2013]

6. Woolf SH, Johnson RE. The break-even point: When medical advances are less important than improving the fidelity with which they are delivered. *Ann Fam Med* 2005;3:545-52

7. Meredith LS, Mendel P, Pearson M, Wu SY, Joyce G, Straus JB, Ryan G, Keeler E, Unutzer J. Implementation and maintenance of quality improvement for treating depression in primary care. *Psychiatric Services* 2006;57:48-55

8. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care* 2003;12:47-52

9. Kelly JA, Spielberg F, McAuliffe TL. Defining, designing, implementing, and evaluating phase 4 HIV prevention effectiveness trials for vulnerable populations. *Jaids-J Acq Imm Def* 2008;47(Suppl 1):S28-S33

10. Journal of Implementation science. About implementation science. http://www.implementationscience.com/about [accessed August 2013]

11. Narayan KMV, Gregg EW, Engelgau MM, Moore B, Thompson TJ, Williamson DF, Vinicor F. Translation research for chronic disease - The case of diabetes. *Diabetes Care* 2000;23:1794-8

12. Newhouse R, Bobay K, Dykes PC, Stevens KR, Titler M. Methodology issues in implementation science. *Medical Care* 2013;51(Suppl 2):S32-40

13. Rycroft-Malone J, Burton CR. Is it time for standards for reporting on research about implementation? *Worldviews on Evidence-Based Nursing* 2011;8:189-90

14. Ogrinc G, Mooney SE, Estrada C, Foster T, Goldmann D, Hall LW, Huizinga MM, LiuSK, MillsP, Neily J, NelsonW, Pronovost PJ, Provost L, Rubenstein LV, Speroff T, Splaine M, Thomson R, Tomolo AM, Watts B. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care* 2008;17(Suppl 1):i13-32

15. Kemple T, Rogers C. A mailed personalised self-management plan improves attendance and increases patients' understanding of asthma. *Prim Care Respir J* 2003;12:110-114

16. Carter EL, Nunlee-Bland G, Callender C. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. *Perspect Health Inf Manage* 2011;8:1b

17. Samann A, Muhlhauser I, Bender R, Kloos C, Muller UA. Glycaemic control and severe hypoglycaemia following training in flexible, intensive insulin therapy to enable dietary freedom in people with type 1 diabetes: a prospective implementation study. *Diabetologia* 2005;48:1965-1970

18. Samann A, Muhlhauser I, Bender R, Hunger-Dathe W, Kloos C, Mullr UA. Flexible intensive insulin therapy in adults with type 1 diabetes and high risk for severe hypoglycemia and diabetic ketoacidosis. *Diabetes Care* 2006;29:2196-2199

19. Sämann A, Lehmann T, Kloos C, Braun A, Hunger-Dathe W, Wolf G, Müller UA. Flexible, intensive insulin therapy and dietary freedom in adolescents and young adults with Type 1 diabetes: a prospective implementation study. *Diabetic Med* 2008;25:592-596

20. Homer CJ, Forbes P, Horvitz L, peterson L, Wypij D, Heinrich P. Impact of a quality improvement program on care and outcomes for children with asthma. *Arch Pediatr Adolesc Med* 2005;159:464-469

21. Gruesser M, Hartmann P, Schlottmann N, Lohmann FW, Sawicki PT, Joergens V. Structured patient education for out-patients with hypertension in general practice: a model project in Germany. *J Human Hypertension* 1997;11:501-506

22. Delaronde S, Peruccio D L, Bauer BJ. Improving asthma treatment in a managed care population. *Am J Managed Care* 2005;11:361-368

23. Bott U, Bott S, Hemmann D, Gerger M. Evaluation of a holistic treatment and teaching programme for patients with Type 1 diabetes who failed to achieve their therapeutic goals under intensified insulin therapy. *Diabetic Med* 2000;17:635-643

24. Bunik M, Federico MJ, Beaty B, Rannie M, Olin JT, Kempe A. Quality improvement for asthma care within a hospital-based teaching clinic. *Academic Pediatr* 2011;11:58-65

25. Haahtela T, Tuomisto LE, Pietinalho A, Klaukka T, Erhola M, Kaila M, Nieminen MM, Kontula E, Laitinen LA. A 10 year asthma programme in Finland: major change for the better. *Thorax* 2006;61:663-670

26. Harris F, Chamings P, Piper S, Levick FR. Development and initiation of a diabetes self-management program for an underserved population. *Diabetes Educator* 2000;26:760-764

27. Learmonth D, Rai S. Taking computerized CBT beyond primary care. *Br J Clin Psychol* 2008;47:111-118

28. Kelly CJG, Booth G. Pharmacist-led structured care for patients with diabetic nephropathy. *Br J Diabetes Vasc Dis* 2008;8:86-88

29. Banister NA, Jastrow ST, Hodges V, Loop R, Gillham MB. Diabetes self-management training program in a community clinic improves patient outcomes at modest cost. *J Am Dietetic Assoc* 2004;104:807-810

30. Hasson H, Blomberg S, Duner A. Fidelity and moderating factors in complex interventions: a case study of a continuum of care program for frail elderly people in health and social care. *Implementation Science* 2012;7:23

31. Hasson H. Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science* 2010;5:67

32. Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? *BMJ* 2004;328;1561-3

33. Perrin KM, Burke SG, O'Connor D, Walby G, Shippey C, Pitt S, McDermott RJ, Forthofer MS. Factors contributing to intervention fidelity in a multi-site chronic disease self-management program. *Implementation Science* 2006;1:26

34. Pinnock H, Adlem L, Gaskin S, Harris J, Snellgrove C, Sheikh A. Accessibility, clinical effectiveness and practice costs of providing a telephone option for routine asthma reviews: Phase IV controlled implementation study. *Br J Gen Pract* 2007;57:714-722

35. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. Lancet 2008;371:1149-50

36. Schulz KF, Altman DG, Moher D, for the CONSORT Group (Consolidated Standards of Reporting Trials). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine* 2010;8:18

37. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. *PLoS Med* 2009;6(6):e1000097.

38. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, for the STROBE initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344-9

39. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney SE, Grp SD. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *Qual Saf Health Care* 2008;17:i3-i9

40. Jarlais DCD, Lyles C, Crepaz N. Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement. *Am J Public Health* 2004;94: 361-366

41. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337;a2390

42. Mohler R, Bartoszek G, Köpke S, Meyer G. Proposed criteria for reporting the development and evaluation of complex interventions in healthcare (CReDECI): guideline development. *Int J Nursing Studies* 2012;49:40-46

43. Benchimol E, Langan S, Guttman A, Moher D, Petersen I, Smeeth L, Sorensen HT, Stanley F, von Elm E. The REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement. Available from http://www.equator-network.org/resource-centre/library-of-health-research-reporting [accessed August 2013]

44. Moher D, Schulz KF, Simera I, Altman DG. Guidance for Developers of Health Research Reporting Guidelines. *PLoS Med* 2010;7:e1000217

45. Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. *Management Science* 1963;9:458-467

46. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Advanced Nursing* 2000;32:1008-1015

47. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus Methods: Characteristics and Guidelines for Use. *Am J Public Health* 1984;74:979-983

48. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Inf Manage* 2004;42:15–29

49. Powell C. The Delphi technique: myths and realities. *J Advanced Nursing* 2003;41:376–382

50. de Meyrick J. The Delphi method and health research. *Health Ed* 2003;103:7-16

51. Murphy MK, Sanderson CFB, Black NA, Askham J, Lamping DL, Marteau T, McKee CM. *Consensus development methods, and their use in clinical guideline development*. HMSO. London 2001

52. Worth A, Nurmatov U, Sheikh A. Key components of anaphylaxis management plans: consensus findings from a national electronic Delphi study. *J R Soc Med Sh Rep* 2010;1:42. DOI 10.1258/shorts.2010.010060

53. Avery AJ, Savelyich BSP, Sheikh A, Cantrill J, Morris CJ, Fernando B, Bainbridge M, Horsfield P, Teasdale A. Identifying and establishing consensus on the most important safety features of GP computer systems: e-Delphi study. *Informatics Prim Care* 2005;13:3–11

54. Sheikh A, Developing consensus on national respiratory research priorities: Key findings from the UK Respiratory Research Collaborative's e-Delphi exercise. *Respir Med* 2008;102:1089-1092

55. Pinnock H, Ostrem A, Rodriguez MR, Ryan D, Stallberg B, Thomas M, Tsiligianni I, Williams S, Yusuf O. Prioritising the respiratory research needs of primary care: the International Primary Care Respiratory Group (IPCRG) e-Delphi exercise. *Prim Care Respir J* 2012;21:19-27