Evaluating the feasibility of complex interventions in mental health services: standardised measure and reporting guidelines

Victoria J. Bird, Clair Le Boutillier, Mary Leamy, Julie Williams, Simon Bradstreet and Mike Slade

Background
The feasibility of implementation is insufficiently considered in clinical guideline development, leading to human and financial resource wastage.

Aims
To develop (a) an empirically based standardised measure of the feasibility of complex interventions for use within mental health services and (b) reporting guidelines to facilitate feasibility assessment.

Method
A focused narrative review of studies assessing implementation blocks and enablers was conducted with thematic analysis and vote counting used to determine candidate items for the measure. Twenty purposively sampled studies (15 trial reports, 5 protocols) were included in the psychometric evaluation, spanning different interventions types. Cohen's kappa (κ) was calculated for interrater reliability and test–retest reliability.

Results
In total, 95 influences on implementation were identified from 299 references. The final measure – Structured Assessment of FEasibility (SAFE) – comprises 16 items rated on a Likert scale. There was excellent interrater ($\kappa = 0.84$, 95% CI 0.79–0.89) and test–retest reliability ($\kappa = 0.89$, 95% CI 0.85–0.93). Cost information and training time were the two influences least likely to be reported in intervention papers. The SAFE reporting guidelines include 16 items organised into three categories (intervention, resource consequences, evaluation).

Conclusions
A novel approach to evaluating interventions, SAFE, supplements efficacy and health economic evidence. The SAFE reporting guidelines will allow feasibility of an intervention to be systematically assessed.

Declaration of interest
V.B. carried out consultancy work for the National Collaborating Centre for Mental Health at the time of the study.

Routine implementation of new technologies and innovation within standard practice is a pertinent issue within healthcare, and one which crosses both geographical and disciplinary boundaries.[1,2] The Cooksey report identified cultural, financial and institutional barriers to the implementation of health research, with recommendations suggesting translational research should be viewed as a key area for future investment.[3] Within England and Wales, policy and treatment decisions are guided by the National Institute for Health and Care Excellence (NICE) guideline programme. Guidelines are typically based on evidence reviews with a focus on efficacy and cost-effectiveness. Likewise, international approaches to quality assurance and evaluation have also aimed to use best available evidence to improve patient care by assisting policy makers and clinicians with the decision-making process.[4] However, implementation of interventions within routine practice often remains low.[5] For example, an audit of four adult community mental health teams within one London trust highlighted that only a minority of eligible patients received the interventions recommended in the schizophrenia guideline update.[6] Recommending interventions that cannot readily be implemented wastes resources.

Feasibility of an intervention is one important characteristic with regard to evidence translation.[7] We define feasibility as the cumulative impact of different influences that have an effect on the implementation of an intervention within a specific healthcare system or practice. Across medical disciplines there is a need to better characterise what is and is not feasible within practice to minimise wasted resources, inform prioritisation decisions and improve effectiveness in health systems. At present no structured and psychometrically validated measure has been specifically designed to assess the feasibility of complex interventions for

implementation within mental health services.[8] Furthermore, despite reporting guidelines such as the CONSORT statement having led to demonstrable improvements in the reporting of studies within high-quality journals,[9] there are no reporting guidelines which allow the feasibility of an intervention to be assessed. This study aims (a) to produce an evidence-based measure of the feasibility of implementing a complex intervention in mental health services within the National Health Service (NHS) and (b) to develop reporting guidelines identifying information to report that allows feasibility to be assessed.

Method

Study design

A focused narrative review was used to inform the development of a measure, the Structured Assessment of FEasibility (SAFE). This was followed by psychometric evaluation and modification of the measure through piloting. Ethical approval was obtained from the South East London Research Ethics Committee 4 (formally known as Joint South London & Maudsley and the Institute of Psychiatry NHS Research Ethics Committee) approval 10/H0807/4.

Literature search

Four data sources were used to identify potential studies for inclusion in the focused narrative review: (a) Google Scholar, NHS evidence and PubMed were searched using the terms "implementation" AND ("barriers" OR "facilitators") AND "mental health"; (b) table of contents for the journal *Implementation Science* from January 1999 until December 2010; (c) hand searching the references of retrieved papers for additional

citations; and (d) recommendations from an implementation science expert.

Eligibility criteria

The review included both quantitative and qualitative papers providing the paper presented factors linked to implementation and met the following inclusion criteria: (a) available in print or downloadable format (PDF file or Word document); (b) focused on mental health or an area directly applicable to mental health such as empowerment or shared decision-making in long-term conditions; (c) the study was either a primary qualitative study with ten or more participants, a quantitative or qualitative survey or systematic review of the literature including either qualitative or quantitative evidence; (d) primary studies were conducted within the UK or (for review studies) a proportion of the included studies were conducted within the UK to ensure applicability to the NHS context; and (e) the study focused on the implementation of a manualised intervention or guideline at the individual staff, team or service level.

Data extraction and tabulation

For each included paper the following data were extracted and recorded in an online database: study methodology, target population, study location, details of the intervention or guideline being implemented and the main implementation barriers and facilitators identified. To assess the quality of the included studies the RATS checklist[10] was used for qualitative papers, the Effective Public Health Practice Project tool[11] was used for quantitative research and the NICE systematic review checklist[12] was used for review studies. For qualitative studies, poor quality was defined as two or more red flags (as indicated on the RATS checklist). Quantitative studies or systematic reviews receiving a negative quality rating on their respective tools were defined as poor quality because for both types of study a negative rating indicates significant evidence of bias within the study. Poor quality studies were excluded.

Development of SAFE

Thematic analysis was used to identify implementation influences – barriers and facilitators, within the included studies. These were tabulated and vote counting used to determine the frequency of each theme across the included papers. Influences included in two or fewer studies were excluded as a result of limited generalisability. The decision to include factors included in three or more papers was a pragmatic decision to reduce the potential number of candidate items. We took this decision to help ensure that the items included in the measure would be generalisable across different interventions and settings within the NHS and not just specific to a particular study. The remaining implementation influences were assessed to check their relevance to characterising the feasibility of an intervention. Only influences that directly related to characteristics of the intervention were included, such as the amount of training required or whether the intervention was manualised for example. Each influence was then operationalised as a single question, for example the implementation barrier lack of time was operationalised as: is the intervention time consuming? Each item was rated as 'yes', 'partial', 'no' or 'unable to rate'. Anchor points for each item were developed based on the consensus opinion of three NHS clinicians and two researchers. The draft measure was then piloted and modified by three members of the research team (one clinician and two researchers) to ensure the rating categories were comprehensively defined and the measure easy to use.

Psychometric evaluation

Within the psychometric evaluation of SAFE, 19 purposively selected papers (reporting on 20 interventions) were rated using the measure (references available from the authors on request). The interventions were described in trial reports ($n = 15$) and study protocols ($n = 5$), and spanned pharmacotherapy ($n = 2$), psychosocial ($n = 12$) and service-based interventions ($n = 6$). To investigate test–retest reliability, each paper was re-rated 1 week later. To investigate interrater reliability, each paper was double rated by at least one of three other researchers. Reliability was measured using weighted Cohen's kappa (κ). Confidence intervals were calculated using Wilson efficient-score method, corrected for continuity with a coefficient > 0.75 representing excellent reliability.[13] Cohen's κ was calculated for overall agreement between raters and to rate agreement by category (yes v. partial v. no v. unable to rate).

Results

Development of the measure

A total of 299 references were identified in the literature search of which 54 articles were potentially relevant and the full text retrieved. Eleven papers were eligible for inclusion.[7,14–23] These comprised four systematic reviews, two narrative reviews, two survey designs and two semi-structured interview studies and one based on expert consensus. Of the 11 papers, 6 assessed facilitators and barriers of implementation within the NHS and 5 reviewed the international literature, including UK-based papers. Additionally, 43 papers were excluded. The most common reason for exclusion was that results of the paper were not applicable to the NHS context (online supplement DS1).

Ninety-five implementation influences (i.e. barriers and facilitators) were identified from the 11 included papers. A total of 39 of these 95 influences related to the characteristics of the intervention so were retained and included in the vote counting (online Appendix DS1).

The most common implementation themes were staff skills required to carry out the intervention, applicability of the intervention to the population of interest and concordance with staff values. From the 39 influences, 18 (listed first in Appendix DS1) were identified in at least 3 papers and were used as candidate items for the measure. Items were then selected through a process of consensus and consultation within the research team, by merging items (e.g. additional skills or knowledge required was merged with the need for additional training), separating items (e.g. cost implications of the intervention was split into cost-effectiveness and the cost of setting up the intervention) and deleting one item (concerning the match with staff values, as this could not be rated based on intervention papers alone). This process produced a 16-item draft measure, comprising eight barriers and eight facilitators of implementation. The measure was piloted and modifications made to the descriptions of each category, including defining the 'unable to rate' category, and adding more detail to items 3 and 14. This resulted in the final measure (available at www.researchintorecovery.com/our-measures).

Both the Cochrane collaboration[24] and the Centre for Reviews and Dissemination guidance[25] recommend against using summary scores on quality assessments to categorise papers within a systematic review, since items within the scale may have unequal weight. Instead, it is recommended that reviewers attend to the individual items of the scale when conducting sensitivity and subgroup analyses. This same approach was therefore adopted for scoring SAFE, whereby the reviewer rates individual items,

without providing an overall summary score, as barriers and facilitators differ in their importance depending on the context.

Psychometric properties

Interrater reliability ($\kappa = 0.84$, 95% CI 0.79–0.89) and test–retest reliability ($\kappa = 0.89$, 95% CI 0.85–0.93) were both excellent. Across all responses, interrater agreement was 89% (95% CI 85–92) and test–retest agreement was 92.5% (95% CI 89–95).

The 'partial' category produced the lowest percentage agreement across different raters and time points (Table 1). Our impression is that the lower consistency was as a result of unclear descriptions given in the papers, rather than because of raters switching to other responses. For example, it was often hard to determine whether an intervention had two or three components or whether the training involved X or Y amount of time. Table 2 provides the frequencies for each response category per item

and suggests the items varied in the proportion of each category response. The overall level of agreement per item (irrespective of response category, for example 'yes', 'no') was consistently very high, ranging from 80 to 100%. Agreement between raters and across time points was 95–100% for over half of the items.

Reporting of implementation influences

The percentage of papers reporting enough information to allow for a rating varied for each item (Table 2). As detailed in Table 3, 90% of papers did not provide enough information for cost saving to be rated, followed by staff training (45%) and ongoing supervision (35%). In contrast, the complexity of the intervention, the applicability of the population, and additional human and material resources were rateable for all papers (i.e. 100%).

Reporting guidelines

Each item from the developed measure was modified and reorganised to produce reporting guidelines (available at www. researchintorecovery.com/our-measures).

Table 1 Percentage agreement for each response category	
Response category	Agreement, % (95% CI)
Interrater	
Yes	84.5 (78.0–89.5)
Partial	57.8 (45.5–69.2)
No	87.0 (76.2–93.5)
Unable to rate	89.4 (76.1–96.0)
Test–retest	
Yes	90.7 (84.9–94.6)
Partial	72.9 (60.7–82.5)
No	89.1 (78.2–95.1)
Unable to rate	85.4 (71.6–93.5)

Discussion

The SAFE scale was developed on the basis of a focused literature review that identified barriers and facilitators of implementation specifically related to characteristics of the intervention being assessed. The resulting tool was demonstrated to be useable across a range of studies from simple pharmacological interventions through to complex service-level innovations, with the psychometric evaluation indicating that SAFE has excellent interrater

Table 2 Number (and percentage) of papers with each response category by SAFE item																
Response	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Yes	6 (30)	16 (80)	6 (30)	2 (10)	4 (20)	5 (25)	7 (35)	1 (5)	20 (100)	10 (50)	17 (85)	10 (50)	1 (5)	13 (65)	18 (90)	19 (95)
Partial	0 (0)	2 (10)	6 (30)	3 (15)	10 (50)	3 (15)	5 (25)	1 (5)	0 (0)	7 (35)	2 (10)	9 (45)	1 (5)	7 (35)	1 (5)	1 (5)
No	5 (25)	2 (10)	3 (15)	8 (40)	5 (25)	12 (60)	4 (20)	18 (90)	0 (0)	3 (15)	1 (5)	0 (0)	0 (0)	0 (0)	1 (5)	0 (0)
Unable to rate	9 (45)	0 (0)	5 (25)	7 (35)	1 (5)	0 (0)	4 (20)	0 (0)	0 (0)	0 (0)	0 (0)	1 (5)	18 (90)	0 (0)	0 (0)	0 (0)

Q, question; SAFE, Structured Assessment of FEasibility.

Table 3 Items able to be rated in the included papers ($n = 20$)			
	n (%)		
	Trial papers (*n* = 15)	Protocol papers (*n* = 5)	Total papers (*n* = 20)
13. Cost saving	2 (13)	0 (0)	2 (10)
1. Staff training	10 (67)	1 (20)	11 (55)
4. Ongoing supervision	10 (67)	3 (60)	13 (65)
3. Time consuming	13 (87)	2 (40)	15 (75)
7. Costly set up	12 (80)	4 (80)	16 (80)
5. Additional human resources	15 (100)	4 (80)	19 (95)
12. Effectiveness	14 (93)	5 (100)	19 (95)
2. Intervention complexity	15 (100)	5 (100)	20 (100)
6. Additional material resources	15 (100)	5 (100)	20 (100)
8. Adverse events	15 (100)	5 (100)	20 (100)
9. Applicable to population of interest	15 (100)	5 (100)	20 (100)
10. Manualised	15 (100)	5 (100)	20 (100)
11. Flexibility	15 (100)	5 (100)	20 (100)
14. Matches prioritised goals	15 (100)	5 (100)	20 (100)
15. Pilotable	15 (100)	5 (100)	20 (100)
16. Reversible	15 (100)	5 (100)	20 (100)

and test–retest reliability. Across the 15 trial reports and 5 trial protocols, frequently unreported aspects included cost information, staff training time and ongoing support and supervision. The SAFE reporting guidelines were developed to identify the information needed in intervention reports that allow SAFE to be rated. We believe that the scale will be useful for three groups. First, for reviewers and policy makers when assessing the evidence base for an intervention. Second, researchers developing an intervention could make use of the scale to ensure they consider factors related to the implementation of that intervention. Finally, the reporting guidelines are intended to be used by authors reporting an intervention.

Strengths and limitations

Although we have demonstrated that SAFE is a useable and reliable measure, our study has a number of limitations. First, the candidate-item selection process was not systematic. Instead we conducted a selective but focused review of the implementation science literature. It is possible that a wider systematic review would have identified additional implementation barriers and facilitators in relation to characteristics of the intervention. Further to this, the review was restricted to mental health services within the NHS. Although this may limit the tool's applicability to other healthcare settings, a number of systematic reviews have identified similar implementation barriers and facilitators in other settings (such as in the USA) and for other long-term health conditions.[17] Furthermore, a number of included reviews assessed the implementation literature on a broader scale. Specifically, for a review to be included in the thematic analysis, it needed to present data that was applicable, but not restricted, to the UK.

A second limitation was the small-scale pilot and psychometric evaluation. Twenty interventions were included in the psychometric evaluation. These were rated by up to four different reviewers, with one reviewer rating each paper a week later to assess test–retest reliability. Although the number of studies was limited, the papers included in the evaluation covered a broad range of interventions (including many featured within NICE clinical guidance). The focus of the psychometric evaluation mirrored the areas important to a systematic review used for evidence appraisal. For example, within good-quality systematic reviews, multiple reviewers will rate included papers (interrater reliability), with the aim of systematic reviews to be reproducible across time (test–retest reliability). The psychometric properties evaluated in this study were selected to reflect these features. Future work could look at evaluating the use of SAFE within an evidence review procedure such as a health technology appraisal or guideline development process.

Finally, the methods used to develop the reporting guidelines were limited in their scope. Moher and colleagues suggest a method for developing reporting guidelines that includes a review of the literature followed by a Delphi exercise and face-to-face consensus meeting.[26] As the reporting guidelines in this study focus specifically on allowing the rating of SAFE within evidence appraisal and decision-making processes, a more pragmatic approach to the development process was undertaken, in that each item in SAFE was constructed as an item in the reporting guidance. Future work could look at expanding these reporting guidelines to include other areas outside of mental health services and implementation features in addition to the characteristics of the intervention.

Despite these limitations, one strength of the study was that the psychometric evaluation indicated that SAFE is useable and reliable. The ease of use of SAFE suggests it could be easily appended to current evidence review processes across a range of different contexts. The associated reporting guidelines also have the potential to have a positive impact on the quality of interventions reported in peer-reviewed journals, thus providing systematic reviewers and policy makers with the information needed to evaluate likely implementation.

Comparison with the literature

Over the past decade implementation science has become a rapidly evolving area of interest with research attention turning to the implementation and sustainability of programmes and innovations within routine clinical care.[27] Within their review of the literature, Wiltsey Stirman and colleagues[28] identified 125 studies investigating sustainability, including 20 studies within the mental health domain. They found that innovation characteristics including fit with current practice, ability for the innovation to be modified and effectiveness were important influences on the sustainability of the innovation being assessed in the individual studies. Furthermore, features such as resources, working culture and training and education requirements also had an impact and match items included in the SAFE scale.

Although SAFE is a novel tool for assessing the feasibility of an intervention at the evidence review stage, other attempts have been made to assess and characterise the barriers to routine translation of evidence into practice. In their review of implementation measures, Chaudoir and colleagues identified 62 available measures assessing different aspects of implementation.[8] None of the identified measures specifically focused on the characteristics of an intervention associated with feasibility, instead the measures were either restricted to evaluations of specific interventions, focused on guideline implementation or including assessment of the innovation alongside other areas such as staff attitudes, political context, organisation factors, all of which would not be possible to assess at the evidence appraisal phase. Furthermore, unlike SAFE, which has demonstrable interrater and test–retest reliability, the majority of measures in the review were not psychometrically evaluated.[8] Although not included in the Chaudoir et al review,[8] the NHS Institute for Innovation and Improvement has recently developed the Spread and Adoption tool, which aims to help staff increase the sustainable implementation of innovations within the NHS.[29] This online-based tool asks individuals to rate their agreement with a number of statements grouped into three categories: people, innovation and context. Although providing a summary assessment, the tool does not specifically focus on rating the feasibility of the intervention and instead covers a broader range of contextual factors; furthermore, it lacks a clear empirical basis. Finally, Slaghuis and colleagues[30] have also developed a framework and instrument to measure the sustainability of new work practices being implemented in long-term care. They identify 'routinisation' and 'institutionalisation' as the two elements of sustainability. However, as with the measures included in the Chaudoir et al review,[8] the framework and measure are designed to evaluate practices within clinical use, rather than at the evidence review stage. By contrast, SAFE assesses individual intervention papers during the policy-making process.

Relevance for practice and policy

To support implementation in clinical practice, an understanding of the factors that facilitate or hinder successful evidence utilisation is required. At present, healthcare improvements have often been targeted at factors related to individual healthcare practitioners, such as their knowledge, routine and attitudes.[27,31] However, successful implementation is influenced by components

occurring at multiple ecological levels of the healthcare system, such as the individual, social, organisation, economic and political context and patient beliefs and behaviour.[7,32–34] Implementation is a complex social process linked with the context in which it takes place.

The SAFE scale specifically focuses on one factor indentified as important to successful implementation, namely the characteristics of the intervention. Within this complex process of implementation, rating feasibility based on the characteristics of that intervention offers a circumscribed and useable source of information for both reviewers and policy makers when making decisions about evidence recommendations. Guideline development processes make use of systematic reviews of best available evidence as part of the decision-making process, alongside other rating systems such as GRADE (Grading of Recommendations Assessment, Development and Evaluation), which makes statements about the overall quality of the evidence. Recently, there have been further suggestions that the GRADE process should incorporate other features of the evidence and intervention including resource allocation.[35] It is at this stage in the evidence review process that SAFE could be used to help clinicians and guideline panellists with the decision-making process.

A number of papers have focused on the implementation of NICE clinical guidelines for mental health conditions. Despite a range of initiatives, implementation within routine care, particularly of psychological therapies and interventions focusing on physical healthcare, has remained low.[14,36,37] For instance, uptake of both family intervention and cognitive–behavioural therapy (CBT) for psychosis has been low, with estimates suggesting that less than 30% of eligible patients receive these interventions.[38] These findings are not restricted to schizophrenia – Rhodes and colleagues[39] found that although the majority of clinicians were aware of and using NICE clinical guidance for depression, only 20% felt confident in their use of the guidelines. Many clinicians stated that resource implications, lack of time and availability of training had a negative impact on their routine utilisation within clinical practice. Using SAFE within the evidence review process could help to highlight areas of interventions that make their implementation more difficult. This would allow for the strategic targeting of resources and the tailoring of implementation strategies at an early stage in the dissemination process to overcome these issues and hence maximise routine implementation. As well as the clinical gains, the cost savings arising from higher levels of implementation are potentially significant. For example, Vos and colleagues[40] indicated that if recommended treatments that are currently underutilised, such as CBT for depression and anxiety and family interventions for schizophrenia, were implemented then significant cost savings would be made, in addition to improvements in the health status of individuals.

The second aim of the paper was to produce a checklist for authors to use when reporting interventions. The pilot study indicated that a number of areas are at present poorly reported in both trial protocols and in randomised controlled trial publications. For instance, despite economic costs and staff time constraints being identified as two main barriers to implementation, few trial publications and protocols reported details of these areas. One way to improve the consistency of reporting within journals is the use of reporting guidelines. Hopewell and colleagues[9] have recently demonstrated that the implementation of CONSORT has led to improvements in the abstracts of articles published in a number of high-quality medical journals. Although the SAFE reporting guidelines have not been developed using a formal framework,[26] they are empirically supported and will support improved characterisation of feasibility.

Future research

Given the interest in implementation science and the increasing evidence to suggest low implementation of evidence within clinical practice, it is imperative that future work continues to assess not only the barriers to implementation but how these can be overcome. The results presented here represent a pilot study and small psychometric evaluation of a new measure and reporting guideline. Larger-scale work is needed to assess the utility of SAFE within systematic reviews such as those used within the guideline development process. Additionally, work could focus on adapting and modifying SAFE so that it is applicable to other areas of healthcare and other non-UK settings. In particular, implementation influences may differ across settings, and the degree of commonality is unknown – future research using the same methodology with different clinical populations and service settings will be needed to establish whether the same influences, and hence SAFE, apply.

Implications

The SAFE scale represents a novel approach to assessing the feasibility of different interventions. It has the potential to be used alongside efficacy and health economic evidence to assist commissioners, policy makers and guideline developers with their decision-making processes. This comes at a time when mental health services worldwide are faced with increasingly difficult decisions regarding resource allocation and implementation priorities. Furthermore, the identification of reporting guidelines for feasibility provides a mechanism for standardising the reporting of this aspect of interventions within high-quality peer-reviewed publications.

Victoria J. Bird, BSc, **Clair Le Boutillier**, MSc, **Mary Leamy**, PhD, **Julie Williams**, MSc, Health Service and Population Research Department, Institute of Psychiatry, King's College London, London; **Simon Bradstreet**, PhD, Scottish Recovery Network, Glasgow; **Mike Slade**, PhD, PsychD, Health Service and Population Research Department, Institute of Psychiatry, King's College London, London, UK

Correspondence: Victoria J. Bird, Health Service and Population Research Department (Box PO29), Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, UK. Email: victoria.bird@kcl.ac.uk

First received 22 Feb 2013, final revision 12 Jun 2013, accepted 5 Sep 2013

Funding

References

1   Tansella M, Thornicroft G. Implementation science: understanding the translation of evidence into practice. *Br J Psychiatry* 2009; **195**: 283–5.

2   McGorry P. Prevention, innovation and implementation science in mental health: the next wave of reform. *Br J Psychiatry* 2013; **202** (suppl 54): s3–4.

3   Cooksey D. *A Review of UK Health Research Funding*. HM Treasury, 2006.

4   Grol R. Has guideline development gone astray? Yes. *BMJ* 2010; **340**: c306.

5   Colom F. Keeping therapies simple: psychoeducation in the prevention of relapse in affective disorders. *Br J Psychiatry* 2011; **198**: 338–40.

6   National Institute for Health and Clinical Excellence. *Schizophrenia: Core Interventions in the Treatment and Mangament of Schizophrenia in Adults in Primary and Secondary Care (Updated Edition)*. NICE, 2009.

7 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. *Implem Sci* 2009; **4**: 50.

8 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. *Implem Sci* 2013; **8**: 22.

9 Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ* 2012; **344**: e4178.

10 Clark J. How to peer review a qualitative manuscript. In *Peer Review in Health Sciences* (eds F Godlee, T Jefferson): 219–35. BMJ Books, 2003.

11 Effective Public Health Practice Project. *Quality Assessment Tool for Quantitative Studies*. EPHPP, 2009 (http://www.ephpp.ca/tools.html).

12 National Institute of Health and Clinical Excellence. *NICE Guidelines Manual*. NICE, 2009.

13 Fleiss JL. *Statistical Methods for Rates and Proportions (2nd edn)*. John Wiley, 1981.

14 Berry K, Haddock G. The implementation of the NICE guidelines for schizophrenia: barriers to the implementation of psychological interventions and recommendations for the future. *Psychol Psychother* 2008; **81**: 419–36.

15 Brooker C, Curran J. The national continuous quality improvement tool for mental health education: results of targeted and supported implementation in England. *J Interprof Care* 2006; **20**: 276–89.

16 Brooker C, Saul C, Robinson J, King J, Dudley M. Is training in psychosocial social interventions worthwhile? Report of a psychosocial intervention trainee follow-up staudy. *Int J Nurs Stud* 2003; **40**: 731–47.

17 Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow practice guidelines? A framework for improvement. *JAMA* 1999; **282**: 1458–65.

18 Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. *Implem Sci* 2007; **2**: 40.

19 Gravel K, Legare F, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: a systematic review of health professionals' perceptions. *Implem Sci* 2006; **1**: 16.

20 Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care* 2005; **14**: 26–33.

21 Michie S, Pilling S, Garety P, Whitty P, Eccles MP, Johnston M, et al. Difficulties implementing a mental health guideline: an exploratory investigation using psychological theory. *Implem Sci* 2007; **2**: 8.

22 Smith G, Velleman R. Maintaining a family work for psychosis service by recognising and addressing barriers to implementation. *J Ment Health* 2002; **11**: 471–9.

23 Williams CHJ. Cognitive behaviour therapy with assertive outreach teams: barriers to implementaton a qualitative peer audit. *J Psychiatr Ment Health Nurs* 2008; **15**: 850–6.

24 Higgins JT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley, 2008.

25 Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare*. Centre for Reviews and Dissemination, 2009.

26 Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010; **7**: e1000217.

27 Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004; **8**(6).

28 Wiltsey Stirman S, Kimberly J, Cook N, Calloway A, Castro F, Charns M. The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. *Implem Sci* 2012; **7**: 17.

29 NHS Institute for Innovation and Improvement. *The Spread and Adoption Tool*. NHS Institute for Innovation and Improvement, 2012 (http://www.institute.nhs.uk/index.php?option=com_spread_and_adoption).

30 Slaghuis SS, Strating MM, Bal RA, Nieboer AP. A framework and a measurement instrument for sustainability of work practices in long-term care. *BMC Health Serv Res* 2011; **11**: 314.

31 Lawrence V, Fossey J, Ballard C, Moniz-Cook E, Murray J. Improving quality of life for people with dementia in care homes: making psychosocial interventions work. *Br J Psychiatry* 2012; **201**: 344–51.

32 Grol R, Wensing M, Eccles M. *Improving Patient Care: The Implementation of Change in Clinical Practice*. Elsevier, 2005.

33 Grol RPTM, Bosch MC, Hulscher MEJL, Eccles MP, Wensing M. Planning and studying improvement in patient care: the use of theoretical perspectives. *Milbank Q* 2007; **85**: 93–138.

34 Sanders D, Haines A. Implementation research is needed to achieve international health goals. *PLoS Med* 2006; **3**: e186.

35 Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, et al. Incorporating considerations of resources use into grading recommendations. *BMJ* 2008; **336**: 1170–3.

36 Gagliardi AR. Translating knowledge to practice: optimizing the use of guidelines. *Epidemiol Psychiatr Sci* 2012; **21**: 231–6.

37 Jolley S, Onwumere J, Kuipers E, Craig T, Moriarty A, Garety P. Increasing access to psychological therapies for people with psychosis: predictors of successful training. *Behav Res Ther* 2012; **50**: 457–62.

38 Prytys M, Garety PA, Jolley S, Onwumere J, Craig T. Implementing the NICE guideline for schizophrenia recommendations for psychological therapies: a qualitative analysis of the attitudes of CMHT staff. *Clin Psychol Psychother* 2011; **18**: 48–59.

39 Rhodes L, Genders R, Owen R, O'Hanlon K, Brown JS. Investigating barriers to implementation of the NICE Guidelines for Depression: a staff survey with community mental health teams. *J Psychiatr Ment Health Nurs* 2010; **17**: 147–51.

40 Vos T, Haby MM, Magnus A, Mihalopoulos C, Andrews G, Carter R. Assessing cost-effectiveness in mental health: helping policy-makers prioritize and plan health services. *Aust N Z J Psychiatry* 2005; **39**: 701–12.

EXTRA CONTENT ONLINE