Cite as

Bird V, Le Boutillier C, Leamy M, Williams J, Bradstreet S, Slade M (2014) *Evaluating the feasibility of complex interventions in mental health services: standardised measure and reporting guidelines*, British Journal of Psychiatry, **204**, 316-321.

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

Victoria J. Bird[1], Clair Le Boutillier[1], Mary Leamy[1], Julie Williams[1], Simon Bradstreet[2], Mike Slade[1].

[1]King's College London, Institute of Psychiatry, Health Service and Population Research Department, (Box 029) De Crespigny park, London, SE5 8AF,

[2]Scottish Recovery Network, Suite 320-321, Baltic Chambers, 50 Wellington Street, Glasgow, G2 6HJ, Simon Bradstreet, Network Director

Corresponding author

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

Abstract

Background: 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 focussed 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 was calculated for inter-rater reliability and test-retest reliability.

Results: 95 influences on implementation were identified from 299 reviewed references. The final measure - Structured Assessment of Feasibility (SAFE) - comprises 16 items rated on a Likert scale. SAFE demonstrated excellent inter-rater (kappa 0.84, 95% CI 0.79 - 0.89) and test re-test 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. SAFE Reporting Guidelines include 16 items organised into 3 categories (Intervention, Resource consequences, Evaluation).

Conclusion: SAFE is a novel approach to evaluating interventions, and supplements efficacy and health economic evidence. SAFE Reporting Guidelines will allow feasibility of an intervention to be systematically assessed.

Introduction

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 Clinical Excellence 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 which cannot readily be implemented wastes resources.

Feasibility of an intervention is one important characteristic in regards to evidence translation.(7) We define feasibility as the cumulative impact of different influences which impact on the implementation of an intervention within a specific health care system or practice. Across medical disciplines there is 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 papers contain enough information to allow the feasibility of an

intervention to be assessed. This study aims (i) to produce an evidence-based measure of the feasibility of implementing a complex intervention in mental health services within the NHS, and (ii) to develop reporting guidelines identifying information to report which allows feasibility to be assessed.

Method

Study design

A focussed narrative review was used to inform the development of a measure. This was followed by psychometric evaluation and modification of the measure through piloting.

Literature search

Four data sources were used to identify potential studies for inclusion in the focussed narrative review:

1) Google Scholar, NHS evidence and PubMed were searched using the terms "implementation" AND ("barriers" OR "facilitators") AND "mental health"
2) Table of contents for the journal Implementation Science from January 1999 until December 2010
3) Hand searching the references of retrieved papers for additional citations
4) 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 10 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) used for quantitative research and the NICE systematic review checklist(12) 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, as for both 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 less studies were excluded due to limited generalisability. The decision to include factors included in two 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 e.g. 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 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 one week later. To investigate inter-rater reliability, each paper was double rated by at least one of three other researchers. Reliability was measured using weighted Cohen's Kappa. Confidence intervals were calculated using Wilson efficient-score method, corrected for continuity with a coefficient >0.75 representing excellent reliability.(13) Cohen's Kappa was calculated for overall agreement between raters and to rate agreement by category (Yes vs. Partial vs. No vs. 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, six assessed facilitators and barriers of implementation within the NHS, and five

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 Data Supplement 1).

Ninety-five implementation influences (i.e. barriers and facilitators) were identified from the 11 included papers. Thirty-nine of these 95 influences related to the characteristics of the intervention so were retained and included in the vote counting (Table 1).

Insert Table 1 here

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, 17 (shown in bold in Table 1) were identified in at least three 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 (Fig 1).

Insert Fig 1 here

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 sub-group 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

Inter-rater 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, inter-rater agreement was 89% (95% CI 0.85 – 0.92) and test re-test agreement was 92.5% (95% CI 0.89 – 0.95).

Insert Table 2 here

The "partial" category produced the lowest percentage agreement across different raters and time points (Table 2). Our impression is that the lower consistency was due to unclear descriptions given in the papers, rather than due to 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 3 provides the frequencies for each response category per item as suggests the items varied in the proportion of each category response. The overall level of agreement per item (irrespective of response category e.g. "yes", "no") was consistently very high ranging from 80-100%. Agreement between raters and across time points was 95-100% for over half of the items.

Insert Table 3 here

Reporting of implementation influences

The percentage of papers reporting enough information to allow for a rating varied for each item (Table 3).

Insert Table 4 here

As detailed in Table 4, 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 (e.g.100%).

Reporting guidelines

Each item from the developed measure was modified and re-organised to produce reporting guidelines (Fig 2).

Insert Fig 2 here

Discussion

The Structured Assessment of Feasibility (SAFE) scale was developed on the basis of a focused literature review which 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 inter-rater and test re-test reliability. Across the 15 trial reports and five trial protocols, frequently un-reported aspects included cost information, staff training time and ongoing support and supervision. SAFE Reporting Guidelines were developed to identify the information needed in intervention reports which 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. Firstly 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 tools applicability to other healthcare settings, a number of systematic reviews have identified similar implementation barriers and facilitators in other settings (such as the US) 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 re-test 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 (inter-rater reliability), with the aim of systematic reviews to be reproducible across time (test re-test 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 (HTA) or guideline development process.

Finally, the methods used to develop the reporting guidelines were limited in their scope. Moher and colleagues suggests a method for developing reporting guidelines which 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 positively 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 last 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. None of the identified measures specifically focussed 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 inter-rater and test re-test reliability, the majority of measures in the review were not psychometrically evaluated.(8) Although not included in the Chaudoir 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. Like many of However like with the measures included in the Chaudoir 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, health care improvements have often been targeted at factors related to individual health care 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 one 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 focussed 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 focussing on physical health care, has remained low.(14, 36, 37) For instance, uptake of both family intervention and cognitive

behavioural therapy 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 which 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 which 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 trial RCT 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 that 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 to be applicable to other areas of healthcare and other non-UK settings. In particular, implementation influences may differ across settings, and 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 Structured Assessment of FEasibility (SAFE) scale represents a novel approach to assessing the feasibility of different interventions. SAFE 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.

Declaration of interest

VB, CL, ML and JW are funded by a National Institute for Health Research (NIHR) Program Grant for Applied Research. MS received a research grant from National Institute for Health Research (NIHR) Program Grant for Applied Research (Grant RP-PG-0707-10040). VB does consultancy for the National Collaborating Centre for Mental Health, no other relationships or activities that could appear to have influenced the submitted work are reported.

Author contributions

All authors made a substantial contribution to the conception and design of the study, data collection and interpretation. All authors were involved in drafting and revising the article and gave their final approval for the version to be published.

Ethical approval

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.

Source of funding

The study is independent research commissioned through a National Institute for Health Research (NIHR) Program Grant for Applied Research (Grant RP-PG-0707-10040) and in relation to the NIHR Specialist Mental Health Biomedical Research Centre at the Institute of Psychiatry, King's College London, and the South London and Maudsley National Health Service (NHS) Foundation Trust. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. The funders did not have a role in the study design, collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

1. Tansella M, Thornicroft G. Implementation science: understanding the translation of evidence into practice. The British journal of psychiatry : the journal of mental science. 2009;195:283 - 5.

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

3. Cooksey D. *A review of UK health research funding*: HM Treasury2006.

4. Grol R. Has guideline development gone astray? Yes. BMJ (Clinical research ed). 2010;340:c306.

5. Colom F. Keeping therapies simple: psychoeducation in the prevention of relapse in affective disorders. The British journal of psychiatry : the journal of mental science. 2011;198(5):338-40.

6. NCCMH. *Core interventions in the treatment and mangament of Schizophrenia in adults in primary and secondary care (updated edition) CG 82*. Leicester: British Psychological Society; 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.

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 (Clinical research ed). 2012;344.

10. Clark J. *How to peer review a qualitative manuscript.* In: Godlee F, Jefferson T, editors. Peer review in health sciences. London: BMJ Books; 2003. p. 219–35.

11. EPHPP. Quality Assessment Tool for Quantitative Studies. EPHPP, . EPHPP2009 [cited 2011]; Available from: (http://www.ephpp.ca/tools.html).

12. NICE. *NICE guidelines Manual*. UK: National Institute of Health and Clinical Excellence; 2009.

13. Fleiss JL. *Statistical methods for rates and proportions.* 2nd ed. New York: 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(3):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, et al. Why don't physicians follow practice guidelines? A framework for improvement. *JAMA*. 1999;282(15):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.

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.

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.

22. Smith G, Velleman R. Maintaining a family work for psychosis service by recognising and addressing barriers to implementation. *J Ment Health*. 2002;11(5):471 - 79.

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* 1ed. UK: Wiley; 2008.

25. CRD. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare*. UK: Centre for Reviews & Dissemination; 2009.

26. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS medicine. 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 Technology Assessment2004.

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. Improvement NIfIa. The spread and adoption tool. NHS Institute for Innovation and Improvement; 2012 [cited 2012 03/12/2012]; Available from: 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 services research. 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. The British journal of psychiatry : the journal of mental science. 2012;201:344-51.

32. Grol R, Wensing M, Eccles M. *Improving patient care: The implementation of change in clinical practice*: Elsevier Ltd; 2005.

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

34. Sanders D, Haines A. Implementation research is needed to achieve international health goals. PLoS medicine. 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 (Clinical research ed). 2008;336:1170-3.

36. Gagliardi AR. Translating knowledge to practice: optimizing the use of guidelines. Epidemiology and psychiatric sciences. 2012 Sep;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. Behaviour research and therapy. 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. Clinical psychology & psychotherapy. 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. The Australian and New Zealand journal of psychiatry. 2005 Aug;39:701-12.

Table 1: Vote counting of the identified influences on implementation (number of influences = 39)

IMPLEMENTATION INFLUENCE	Identified in N papers (%)
Staff skills to deliver the intervention	9 (82)
Applicability of the intervention (to Service users)	8 (73)
Match with staff values, attitudes – does it clash with preferred treatment approach and culture of the team, staff preference?	8 (73)
Staff knowledge to deliver the intervention	7 (64)
Time constraints	7 (64)
Ongoing support and supervision	5 (45)
Outcome expectancy (efficacy) – Do staff think the intervention will work? Etc.	5 (45)
Cost-benefit of intervention (financial)	5 (45)
Cost-benefit (efficacy, risk etc.) – perception of advantage, risks, regret for doing or not doing the intervention	5 (45)
Match with the organisational culture – does it link with values, attitudes of the organisation, is it supported etc.	4 (36)
Match with current practice - Is the intervention breaking routines and habits? Are there contradictory practices or guidelines. Conflict with usual routines and roles	4 (36)
Lack of resources	4 (36)
Flexibility / modifiability – can the intervention be adapted to fit the local context and situation	4 (36)
Guideline / intervention availability including availability of a manual or guide	3 (27)
Confidence in the intervention – lack of confidence in the developer, approach, evidence-based, credibility of the intervention and source.	3 (27)
Lack of reimbursement or incentives to do the intervention	3 (27)
Complexity of the intervention – is the intervention simple or complex	3 (27)
Reversibility and trialability - are the changes permanent or can they be trialled	3 (27)
Service user involvement including in the design of the intervention	2 (18)
Outcome expectancy (observability) – time needed before the results become apparent, are the results observable	2 (18)
Role match – does the intervention challenge the social roles and professional identity of staff.	2 (18)
Intervention is too rigid, cook book and biased	2 (18)
The intervention challenges staff autonomy	2 (18)
Quality of design of the intervention	2 (18)
Degree to which the action done by the team, organisation or individual is disruptive or radical	2 (18)
Stressful nature of the intervention	2 (18)
Time needed to keep up to date with the intervention	1 (9)
Is the source of the intervention internal or external to the organisation	1 (9)
Forgetting the intervention (content) – forgetting the content of the intervention	1 (9)
Forgetting the intervention (action) – forgetting to do the intervention	1 (9)
Divisibility – being able to separate out components of the intervention to implement at different times	1 (9)
Centrality – does the intervention effect a central or peripheral activity	1 (9)
Duration of change and how long will it take	1 (9)
How much attention does the intervention require	1 (9)
Will staff observe others doing the intervention	1 (9)
Lack of trained supervisors	1 (9)
Lack of opportunities for co-working	1 (9)
Adaption of the intervention for sensory impaired groups	1 (9)
Does the intervention allow for patient preference	1 (9)

Table 2: Percentage agreement for each response category

Response category	Agreement % (95% CI)
INTER-RATER	
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 RE-TEST	
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)

Table 3: Number (and percentage) of papers with each response category by SAFE item

	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)

Table 4: Items able to be rated in the included papers (n=20)

Item	Trial papers (n=15) n (%)	Protocol papers (n=5) N (%)	Total papers (n=20) n (%)
13 – Cost saving	2 (13.3)	0 (0)	2(10)
1 – Staff training	10 (67)	1 (20)	11 (55)
4 - Ongoing supervision	10(66.7)	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 (93.3)	5(100)	19 (95)
2 – Intervention complexity	15 (100)	5 (100)	20 (100)
6 – Additional material resources	15 (100) (20	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)

Figure 1: SAFE Scale Version 1

Structured Assessment of FEasibility (SAFE) Scale Version 1

SAFE assesses the extent to an intervention is feasible for implementation in mental health services in the National Health Service (NHS) in England.

BLOCKS Sub-scale
*These items are blocks to implementation. Circle **one** answer for each question.*

1. **Do staff require specific training to deliver the intervention?**

Yes Partial No Unable to rate

***Yes**: The intervention requires more than four hours of training*
***Partial**: The intervention requires up to four hours of training*
***No**: The intervention does not require any specific training*
***Unable to rate:** Not enough information provided to rate item*

2. **Is the intervention complex?**

Yes Partial No Unable to rate

***Yes:** The intervention is made up of more than three separate components*
***Partial:** The intervention contains two or three separate components*
***No:** The intervention only has one component*
***Unable to rate:** Not enough information provided to rate item etc*

3. **Is the intervention time consuming to provide?**

Yes Partial No Unable to rate

***Yes:** The intervention requires more than two hours per week of work (per client)*
***Partial:** The intervention requires more half an hour but less than two hours or work per week (per client)*
***No:** The intervention requires less than half an hour per week (per client)*
***Unable to rate:** Not enough information provided to rate item*

4. **Does the intervention include/require ongoing support and supervision?**

Yes Partial No Unable to rate

***Yes:** The intervention requires an extra weekly supervision or support session*
***Partial:** The intervention requires an additional monthly supervision or support session*
***No:** The intervention does not require any additional support sessions or supervision*
***Unable to rate:** Not enough information provided to rate item*

5. **Does the intervention require additional human resources?**

Yes Partial No Unable to rate

***Yes:** The whole team is required to provide the intervention.*

Partial: *More than one member of staff are involved in providing the intervention*
No: *The intervention can be provided by one member of staff*
Unable to rate: *Not enough information provided to rate item*

6. **Does the intervention require additional material resources?**

Yes Partial No Unable to rate

Yes: *Sizeable resources or special equipment which staff would not usually have access to e.g. a dedicated room, instruments, art materials.*
Partial: *The intervention requires additional but readily available resources e.g. computers*
No: *The intervention does not require any additional resources that staff would not usually have access to.*
Unable to rate: *Not enough information provided to rate item*

7. **Is the intervention costly to set up?**

Yes Partial No Unable to rate

Yes: *The intervention is likely to be too costly to provide without extra funding*
Partial: *The intervention is likely to require other costs to be de-prioritised*
No: *The intervention cost is low*
Unable to rate: *Not enough information provided to rate item*

8. **Are there known adverse events associated with the intervention?**

Yes Partial No Unable to rate

Yes: *There are known serious adverse events associated with the intervention*
Partial: *There are known adverse events associated with the intervention*
No: *There are no known serious or adverse events associated with the intervention*
Unable to rate: *Not enough information provided to rate item*

ENABLERS Sub-scale
*These items are enablers of implementation. Circle **one** answer for each question.*

9. **Is the intervention applicable to the population of interest (e.g. adults using community mental health teams)**

Yes Partial No Unable to rate

Yes: *The intervention has been designed for the population of interest*
Partial: *The intervention has been designed for a general mental health population or can be adapted to be applicable to the population of interest*
No: *The intervention is not applicable to the population of interest*
Unable to rate: *Not enough information provided to rate item*

10. **Is the intervention manualised?**

Yes Partial No Unable to rate

Yes: *All aspects of the intervention are manualised*
Partial: *Some components of the intervention are manualised*

No: *The intervention is not manualised*
Unable to rate: *Not enough information provided to rate item*

11. **Is the intervention flexible (i.e. can be tailored to the context and situation)?**

Yes Partial No Unable to rate

Yes: *The intervention is flexible and can be tailored to the context and situation*
Partial: *Elements of the intervention can be tailored to the context and situation*
No: *the intervention cannot be tailored to the specific context*
Unable to rate: *Not enough information provided to rate item*

12. **Is the intervention likely to be effective (i.e. evidence based and expected to produce positive outcomes)?**

Yes ***Partial*** ***No*** ***Unable to rate***

Yes: *There is an established evidence base regarding the effectiveness of the intervention (e.g. clinical trials)*
Partial: *There is some evidence for the effectiveness of the intervention (e.g. case studies but no clinical trials)*
No: *There is no evidence base for the intervention*
Unable to rate: *Not enough information provided to rate item*

13. **Is the intervention cost saving?**

Yes Partial No Unable to rate

Yes: *The intervention has been demonstrated to save costs*
Partial: *The intervention has been demonstrated to be cost neutral and/or cost effective*
No: *The intervention is not cost saving or cost effective*
Unable to rate: *Not enough information provided to rate item*

14. **Do the intended goals of the intervention match the prioritised goals of the NHS?**

Yes Partial No Unable to rate

Yes: *The primary aims of the intervention match values NHS outcomes e.g. improving mental health and wellbeing, supporting clinical and personal recovery, promoting good physical health, improving service satisfaction, reducing stigma and discrimination [Taken from No Health Without Mental Health, 2011, Department of Health]*
Partial: *The secondary aims of the intervention match the current valued outcomes*
No: *The primary and secondary aims of the intervention do not match the current valued outcomes of the NHS*
Unable to rate: *Not enough information provided to rate item*

15. **Can the intervention be piloted?**

Yes Partial No Unable to rate

Yes: *The intervention can be piloted by a few members of staff AND with only a few service users.*
Partial: *the intervention can be piloted by a few members of staff OR with a few*

service users
No: *The intervention cannot be piloted*
Unable to rate: *Not enough information provided to rate item*

16. **Is the intervention reversible?**

Yes Partial No Unable to rate

Yes: *It is possible to stop the intervention without harmful, or unwanted, effects*
Partial: *It is possible to stop the intervention, but there are likely to be some harmful, or unwanted, effects*
No: *It is not possible to stop the intervention without serious adverse effects*
Unable to rate: *Not enough information provided to rate item*

Further information and downloadable information at researchintorecovery.com/safe.

Figure 2 – SAFE Reporting Guidelines for characterising implementation influences

These reporting guidelines identify the characteristics of a mental health intervention to report in order to allow influences on implementation within the NHS to be evaluated.

Intervention

Item 1: Details of the intervention components

Descriptor: The complexity of the intervention should be specified, this includes recording and listing how many separate components make up the intervention.

Item 2: Intervention Manual

Descriptor: Is the intervention manualised? The report should contain details of any intervention manuals developed or used.

Item 3: Flexibility

Descriptor: Can the intervention be tailored to different contexts and environments?

Item 4: Ability to Pilot the intervention

Descriptor: Can the intervention be piloted with a few individuals or within one or two teams?

Item 5: Reversibility

Descriptor: Are the effects of the intervention permanent or can the intervention be stopped at any point within any harmful effects. If there are likely to be adverse effects associated with discontinuing the intervention, these should be reported.

Item 6: Population

Descriptor: The intended population of the intervention should be described. For example is the intervention aimed at people with a particular diagnosis or using a particular service? The ability to adapt the intervention for use within other populations should also be reported.

Resource consequences

Item 7: Staff training

Descriptor: Do staff require any specific training to deliver the intervention? If yes, details of the training should be reported. This includes the name of any specific training, the length of training e.g. does it last two half days, three hours etc. and any details about booster training sessions.

Item 8: Support and supervision

Descriptor: Any ongoing support and supervision required to deliver the intervention should be reported. This included details about how much supervision is recommended and the format of supervision, e.g. Individual, group, peer supervision etc.

Item 9: Time costs

Descriptor: How much time does the intervention require per client per week?

Item 10: Human resources

Descriptor: what human resources are required to deliver the intervention?

Item 11: Material resources

Descriptor: what material resources are required to deliver the intervention?

Item 12: Set-up costs

Descriptor: Where possible the cost implications of the intervention should be reported. Any estimated costs associated with setting up the intervention should be reported.

Evaluation

Item 13: Efficacy

Descriptor: This relates to the existing evidence base for the intervention or any theoretical evidence base. For instance is there supporting evidence that the intervention is efficacy of the intervention? Has the effectiveness of the intervention been established in previous clinical trials?

Item 14: Outcomes

Descriptor: what are the intended outcomes of the intervention? What are the primary outcomes? What are the secondary outcomes?

Item 15: Cost saving

Descriptor: Any information relating to the costs of the intervention should be reported, including the potential costs saved. Is there evidence of cost saving? Has the cost effectiveness of the intervention been assessed or estimated?

Item 16: Adverse events

Descriptor: Are there any known adverse events associated with the intervention? What adverse events might be anticipated?