IDEAS
Evidence to improve maternal & newborn health

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Measuring implementation strength

Literature review draft report 2012

Acknowledgements

Dr Joanna Schellenberg, Principal Investigator, IDEAS project, London School of Hygiene & Tropical Medicine.

Funded by the Bill & Melinda Gates foundation

Review supervised by Dr Bilal Iqbal Avan
Review prepared by Dr Natalia Bobrova
Copy edited by Kate Sabot
Coordination of publication by Agnes Becker

ISBN - 978 0 902657 90 9

w: ideas.lshtm.ac.uk

IDEAS
Evidence to improve maternal & newborn health

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Contents

Acronyms and abbreviations

ACIC	Assessment of Chronic Illness Care
ACT	Assertive Community Treatment
AIDS	Acquired Immune Deficiency Syndrome
BSC	Balanced Score Card
CCC	Community Care Centres
CCM	Chronic Care Model
CQI	Continuous Quality Improvement
DACT	Dartmouth Assertive Community Treatment
HIV	Human Immunodeficiency Virus
IDEAS	Informed Decisions for Actions
IPS	Individual Placement Support
IRT	Implementation Rating Instrument
ISD	Integrated Delivery System
MHI	Medical Home Index
NACO	National AIDS Control Organization
NGOs	Non-Governmental Organisations
PDC	Project Developmental Continuity
PRISMA	Providing Innovative Service Models and Assessment
QI	Quality Improvement
QRF	Quarterly Report Form
RED	Reaching Every District approach
SAT	System Assessment Tool
SE	Supported Employment
TB	Tuberculosis
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNICEF	United Nations Children's Fund
WHO	World Health Organization

List of tables

List of figures

Introduction

Measuring implementation strength (sometimes referred to as implementation intensity) is an important programme evaluation process which helps to understand why some programmes are successful and some fail, attribute outcomes to a programme, and anticipate outcomes of future programmes (Bryce et al., 2011, Dane and Schneider, 1998, Durlak, 1998b, Dusenbury et al., 2003, Hall et al., 1994, Proctor et al., 2011, Victora et al., 2011). Implementation data can also help in evaluating and improving progress toward specific outcomes and intervention strategies (Damschroder and Hagedorn, 2011, Durlak, 1998b, McGraw et al., 2000, Proctor et al., 2011, Mowbray et al., 2003). In developing a new approach, the 'District Evaluation Platform', to evaluate large-scale effectiveness for proven interventions at a national level, Victora et al., emphasise the importance of measuring implementation strength as "insufficient implementation is a common reason for absence of impact" (Victora et al., 2011). Nevertheless, despite the importance of evaluating implementation strength of complex, multidimensional interventions, scientific evidence devoted to this issue is limited, especially in low income countries.

The common evaluation framework (Figure 1. Monitoring and evaluation of health systems strengthening. (WHO, 2010)), initially developed to measure health system performance and recently used in measuring implementation strength of community case management of childhood illness in six African countries, identifies four major indicators in measuring health systems: 1) inputs & processes, 2) outputs, 3) outcomes, and 4) impact (Bryce et al., 2011, Victora et al., 2011, WHO et al., 2010). The main aim of the review was to synthesise existing literature on quantifiable measurements of implementation strength and address the following specific questions:

Primary questions

- What are the components of programme implementation and how are they categorised?

  The identification of essential programme components and their categorisation is one of the most important tasks in developing tools for implementation evaluation of large-scale, complex interventions. Clear definitions of components to be measured are paramount in quantifying implementation of programmes in different contexts.

- How is implementation strength measured?

  The review focused only on methods for quantifiably measuring implementation such as scoring or rating systems: how scores were calculated or scales built, and how composite implementation strength scores were calculated.

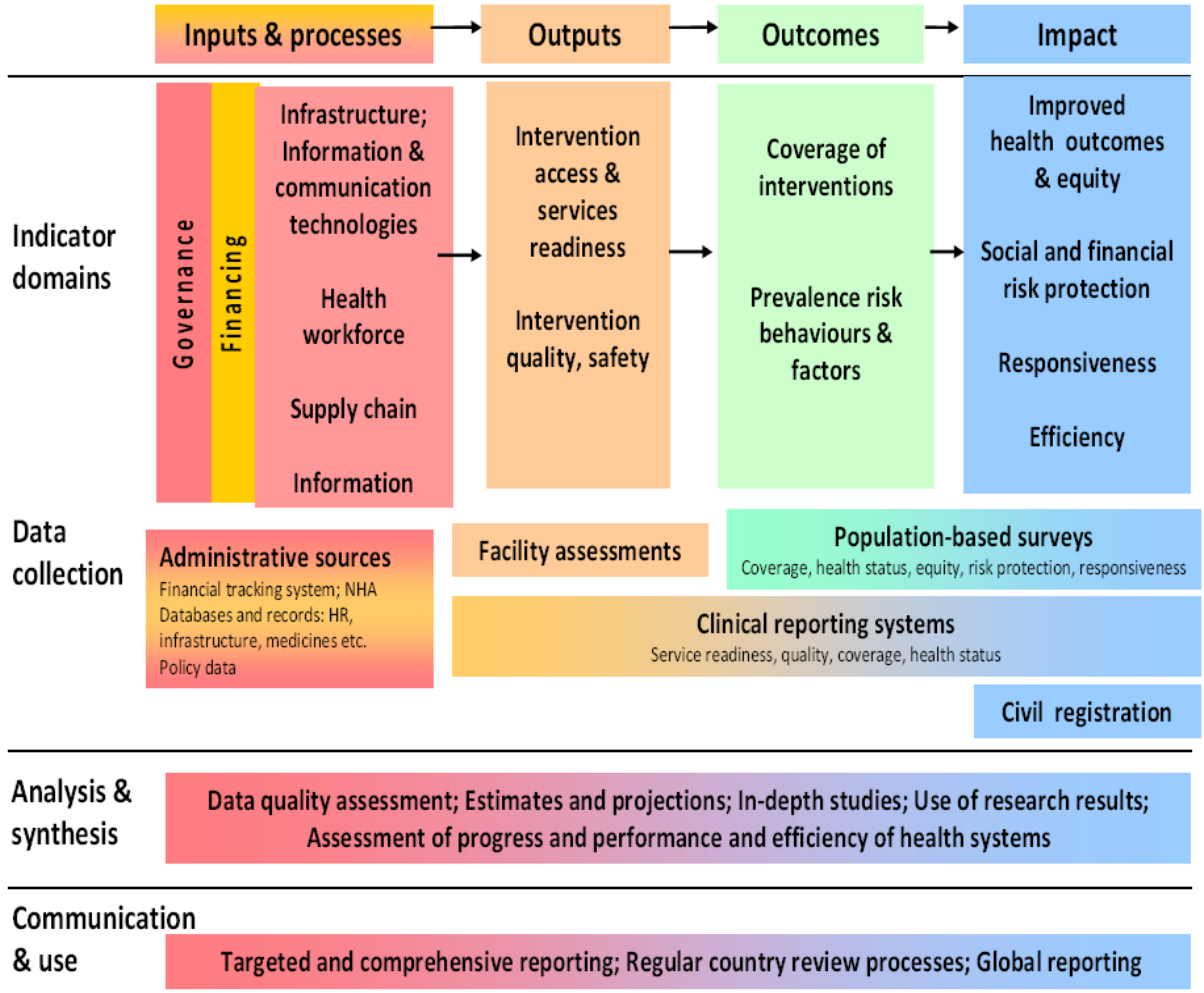
Secondary questions

- Do examples exist of evaluations of diverse programmes, with different inputs, process and outputs all aiming for a common outcome, where a single implementation strength score have been used?

- Does evidence exist that interventions with higher implementation strength produce better outcomes? Are there examples of studies that compare effects of implementation with different strength level?

A particular focus of the review was on studies and reports which used or developed composite scores, that is, multidimensional measures of implementation strength of complex interventions.

Figure 1. Monitoring and evaluation of health systems strengthening. (WHO, 2010)



Methods

Given the scarcity of literature on measuring implementation strength and the absence of an agreed set of terms in the implementation evaluation field, the review is descriptive, rather than exhaustive. Peer reviewed and grey literature was searched on a variety of relevant subjects, and was not restricted to the field of maternal, newborn or public health.

Inclusion criteria

The review was focused on the *methods* used to measure implementation strength rather than on the *results* of any particular study. Thus, the review was not restricted to any specific study design (e.g., randomised controlled trials, cross-sectional, time-series), type of participant or type of intervention or study field (e.g., literature from education, mental health, substance abuse, behavioural, nursing, business fields are all included). Nevertheless, included studies had to contain detailed descriptions of methodology and tools used to measure implementation. In this sense, the major screening criterion was whether the paper provided detailed process information for measuring implementation, such as tool development, scales, and scores of a given construct, psychometric characteristics of development or validation. All in all, selected papers had to:

1) Describe methodologies for measurement of implementation in detail;
2) Report on the development of these methodologies.

The second criterion was useful in circumstances where multiple studies used the same tool or method. In these cases the study with the richest description of the process of developing the methodology was included. This approach was adopted as the purpose of this review is to provide an understanding of the range of tools and methodologies in use for implementation strength.

Exclusion criteria

The review was restricted to studies which involved human subjects and communicated in English.

The search strategy

The following databases were used for the peer reviewed articles:

- MEDLINE, (1970 to June 30, 2012)
- EMBASE, (1970 to June 30, 2012)
- ERIC, (1970 to June 30, 2012)
- Web of Science, (1970 to June 30, 2012)

Search terms

The search had three major components which were entered as two words phrases: 1) implementation, 2) measure* or evaluate* or assess* or examine*, 3) strength or intensity or extent or degree or rate or scale. The detailed search strategy is provided in Appendix 1, Table 4. Search strategyThe results from the searches were downloaded to the EndNote program where duplicates were eliminated. Titles, abstracts, full texts and reference lists and bibliographies were hand searched. Data from the selected articles were extracted into a specially developed form (Appendix 1, Table 7. Data extraction form for the tools measuring implementation strength.). Grey literature

was searched through websites (Appendix 1, Table 5. Main websites used in grey literature search.) using keywords "measuring/evaluating/assessing implementation", "implementation research", "implementation intensity/strength/degree/level/extent/rate". Experts' suggestions were used to identify relevant peer reviewed articles and grey literature papers. (See expert list: Appendix 1, Table 6. List of experts contacted.). The search strategy also included hand searching the following journals: *Implementation Science, Evaluation Review, International Journal of Quality of Health Care, McKinsey Quarterly*.

Data extraction

The data was extracted by three independent researchers into a data extraction form (Appendix 1, Table 7. Data extraction form for the tools measuring implementation strength.), developed and reviewed by the review authors, including the following sections:

1. Author, year of the study
2. Study context/setting
3. Background
4. Conceptual basis
5. Methodological development steps
6. Psychometric characteristics
7. Structure
8. Contents
9. Scaling
10. Scoring
11. Data source
12. Conduct
13. Association information
14. Evidence of further use of the tool

Data synthesis

The findings were synthesised and presented in a manner aligned with the objectives of the review. The selected studies were presented using narrative synthesis techniques such as exploring relationships within and between studies, and providing case descriptions (e.g., description of tools and methodologies). In addition, a glossary of definitions of implementation strength and measured elements of implementation was developed (e.g., implementation quantity, dose, intensity).

With the described methodology 2,297 titles and abstracts were identified in total. After reviewing the full texts of 184 articles that initially met the inclusion criteria, 26 studies were selected for review synthesis (Table 1: Included study characteristics).

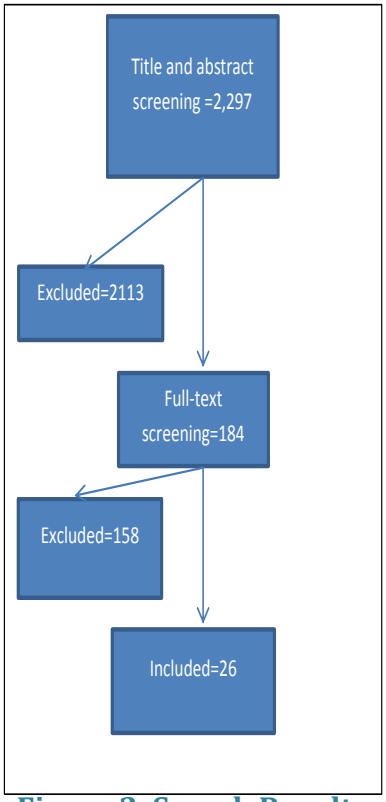


Figure 2. Search Results

Limitations

The review search criteria focused on measurements of implementation and could have missed the studies related to similar topics such as diffusion of innovations. Classroom based studies and studies using observations, video-recording, and expert rating were not included as they were found less useful for the "District Evaluation Platform" evaluation approach.

Results and discussion

In this section an overview of the results is provided, followed by a breakdown of the results by the study objectives.

I. Results overview

The review included studies from mental health, chronic care, primary care, public service, health promotion, public health, and education fields. Sixty-two percent of the studies were based in the USA and Canada. All articles, except one, were peer-reviewed (Appendix 1, Table 2). Most of the selected articles investigated implementation in different geographies, communities and settings (d'Abbs et al., 2008, Glanz et al., 1992, Gold et al., 1993, Grizzard et al., 2006, Hacker and Washington, 2004, Hebert and Veil, 2004, Kansanaho et al., 2005, Lee et al., 2002, Orwin, 2000, Paulsell et al., 2002, Pearson et al., 2005, Perez et al., 2011, Ryman et al., 2011, Shortell et al., 1995, Sogarwal and Bachani, 2011, Teague et al., 1998, Vinson et al., 2001); and several articles looked at different time points (Edward et al., 2011, Hebert and Veil, 2004, Paulsell et al., 2002, Ryman et al., 2011, Vinson et al., 2001). Eleven studies conducted their evaluations at a national level, twelve at a regional level, one study compared implementation in two different communities, and two studies tested implementation evaluation framework models and developed case-studies for the models. Two studies investigated programme implementation in rural communities. Target populations of the programmes evaluated included the general population, young children, women and newborns, people with chronic illness, indigenous people, elderly, people with mental illness, homeless population, and people with HIV/AIDS (Appendix 1, Table 2. Data extraction for measurements of implementation.).

In the majority of the studies, implementation quality (studies often use the terms fidelity or integrity) and quantity were measured. In some studies, implementation quantity or implementation 'dosage' was differentiated as 'dose' delivered and 'dose received' (Hebert and Veil, 2004, McGraw et al., 2000, Rosecrans et al., 2008, Wilson et al., 2010). Rosecrans et al., for example, defined implementation quantity as "'dose delivered'- the number of units delivered by interventionists; 'dose received' - the extent to which the target audience actively engages in and receives intervention activities; 'reach'- the amount of target audience that participates in the intervention". Implementation quality was often defined as the extent to which delivery of an intervention adheres to the protocol, guidelines, programme or treatment model originally developed (Bond et al., 1997, Dusenbury et al., 2003, Falloon et al., 2005, O'Donnell, 2008, Rubin et al., 1982, Scheirer et al., 1995, Teague et al., 1998, Wickizer et al., 1998). Only one study has used the exact term "implementation strength" in its tool development, and defined it as "the aggressiveness of the States' efforts to deliver the programme" (Gold et al., 1993). Appendix 1, Table 3. Glossary of implementation strength. provides a glossary of terms related to measurements of implementation such as implementation strength, intensity, implementation degree, implementation level and extent.

Table 1: Included study characteristics

Authors	Year	Country	Setting	Discipline
Smith et al	1977	USA	Pre-school/elementary school	Education
Rubin et al	1982	USA	Community-based education programme	Education
Glanz et al	1992	USA	Physician offices	Health
Gold et al	1993	USA	State MCH programmes	Health
Shortell et al	1995	USA	Hospital quality improvement programmes	Health
Bond et al	1997	USA	Mental illness programmes	Health
Teague et al	1998	USA	Community-based mental illness programme	Health
Orwin et al	2000	USA	Community-based homeless substance abuse programmes	Health
Vinson et al	2001	USA	Community-based children's mental health programme	Health
Bonomi et al	2002	USA	Facility-based chronic disease management	Health
Lee et al	2002	Korea	Hospital quality improvement programmes	Health
Pausell et al	2002	USA	Early childhood development programme	Health
Cooley et al	2003	USA	Special needs children health programmes	Health
Hacker & Washington	2004	Bostwana	Public service organisational change/performance management system	Civil Service
Herbert et al	2004	Canada	Elder care programme	Health
Kanasaharo et al	2005	Finland	Community-based pharmacy counselling	Health
Pearson et al	2005	USA	Facility-based chronic disease management	Health
Grizzard et al	2006	USA	Maternity hospital breastfeeding guidelines	Health
D'Abbs et al	2008	Australia	Facility-based indigenous community chronic disease management	Health
Wilson et al	2010	USA/Canada	Worksite physical activity programme	Health
Perez et al	2011	Cuba	Community-based dengue fever programme	Health
Ryman et al	2011	North Sudan	'Reaching Every District' Immunisation programme	Health
Sogarwal & Bachani	2011	India	Community-based HIV services	Health
Yumo et al	2011	Cameroon	Facility-based TB/HIV activities	Health

II. What are the components and how are they categorised?

It is recognised by many researchers that the identification of critical components or dimensions of programmes and interventions is a crucial step in implementation evaluation (Fixen et al., 2005, Proctor et al., 2005, Mowbray et al., 2003, Teague et al., 1998, Scheirer et al., Rubin, 1982). In fact, it was argued by some implementation scale developers that to apply any model on a wider scale and relate programme contribution to its components' effectiveness, it is necessary to:

- Identify "very explicit operational criteria for programme structure and processes";
- Provide "careful measurement of the multiple dimensions of [the] complex model" (Teague et al., 1998); and
- Understand "the core intervention components may allow for more efficient and cost effective implementation, and lead to confident decisions about what can be adapted to suit local conditions at an implementation site" (Fixsen et al., 2005).

Apart from three studies all provided a conceptual basis for tool development and component identification. These descriptions often included evaluation theories, quality improvement models, and specific programmes models. In the reviewed studies the component identification and grouping were conducted using different sources of information. In the majority of cases, components were identified by researchers and experts who used models or programme guidelines, literature searches on a relevant models or treatment modules, and field implementation experiences. For example, Teague et al., based key dimensions of programme implementation on the Assertive Community Treatment (ACT) model, literature describing the model, results from previous work on the model implementation and expert opinion (Teague et al., 1998). The authors further refine the items using suggestions from the colleagues working on the model with final editorial provided by independent evaluators. Bonomi et al., based their tool content on specific interventions and concepts of the Chronic Care Model, as well as by using expert panels (Bonomi et al., 2002). Similarly, the identification and grouping of the indicators under domains and/or sub-domains followed specific models, programme requirements, protocols or guidelines. For example, Grizzard et al., grouped their items according to the UNICEF designed "Ten Steps of Breastfeeding" model (Grizzard et al., 2006); Lee et al., and Shortell et al., used the adapted standard for organisational excellence from the Malcolm Baldrige National Quality Award Criteria (Lee et al., 2002, Shortell et al., 1995, U.S. Chamber of Commerce, 1993); Edward et al., adopted the Balanced Score Card tool for their health systems performance evaluation(Edward et al., 2011).

The content of components identified varied from one study to the next, but often included elements of structure/infrastructure (the framework, organisational arrangements of service delivery) and processes of care (how services are delivered) (Mowbray et al., 2003, Vinson et al., 2001). In many cases, program content included elements of: policies; leadership and coordination; staff structure, training and functioning; type of services/activities provided (structure, nature, quality); resources; community linkages; data management and information systems. Several studies included in their component lists the number or percentage of their target populations, and/or reach of target populations and subpopulations, e.g., "the extent to which the clientele using the services corresponds to the clientele initially targeted"(Gold et al., 1993, Hebert and Veil, 2004, Orwin, 2000, Yumo et al., 2011), staff and patients' perceptions/satisfaction (Edward et al., 2011, Glanz et al., 1992), and financing (Edward et al., 2011, Gold et al., 1993, Vinson et al., 2001).

Where some studies only provided the availability and/or quantity of a particular item (e.g., availability of vaccination kits or percentage of manuals distributed[Yes/No]), others provided more detailed quantitative and descriptive information of their essential elements, e.g., Rubin et al., provided a criterion, or target level for 'home visits' programme activity: *50% of paraprofessional's employment time [spent conducting home visits], at least 75% of the scheduled home visits will be completed for at least 80% of the programme children."* (Rubin et al., 1982).

The detailed contents of each study is provided in Appendix 3. Studies' descriptions.

III. How is implementation strength measured?

The majority of the studies provided methodological development steps (Appendix 1, Table 2. Data extraction for measurements of implementation.) which often included identifying essential components to be measured, grouping components into domains (thematic areas), building a measurement instrument, piloting the instrument, and finalising the instrument. In some articles the development of frameworks was described, levels were developed and specified for each component (on which basis the components would receive one or another rating), and weights were assigned for the components and domains. Only two studies used secondary data alongside the primary data for their tools development. In 14 papers structured and unstructured tools were used, 10 papers described structured instruments, and two used semi-structured instruments.

Most studies used scales and/or scoring systems to measure implementation. Implementation scales included:

1) Scales with detailed descriptions, 'anchors' for each criteria and level

(Bonomi et al., 2002, Cooley et al., 2003, d'Abbs et al., 2008, Rubin et al., 1982, Teague et al., 1998).

These scales were built in a way that not only provides information as to the extent or effort with which the criteria were implemented (e.g., low/medium/high level) but also contains a detailed description and specifications of each level for each attribute which makes a judgement more objective and measurable (Rubin et al., 1982).

Example: Teague et al. 1995, describe a low level of implementation in 'intensity of service' as "fifteen minutes of contact per week or less", and high as "two hours of contact per week or more"(Teague et al., 1995). In this study, anchors were developed for each quantitative and qualitative item endpoint (on a 1-5 scale).

2) Percentage scales

(Edward et al., 2011, Gold et al., 1993, Grizzard et al., 2006, Hebert and Veil, 2004, Rosecrans et al., 2008, Sogarwal and Bachani, 2011).

Percentage scales or percentage scoring systems imply components of programme implementation measured on 0-100 scale (either as percentages, or 0-100 rating points which get converted into percentages). In some studies points for items and/or domains are weighted cumulatively producing a score of 100 (Gold et al., 1993, Hebert and Veil, 2004).

Example: Hebert, et al., measured the degree of implementation by weighting the domains and its indicators and scoring them on a 0-100 percentage scale. The authors concluded that Integrated Service Delivery system was implemented at the rate of 73%, 71% and 70% (Hebert

and Veil, 2004). Grizzard et al., created a two-stage algorithm converting yes/no answers and 5-point scale answers to 0-100% overall scale for each domain, and further categorised the scores into descriptive low/partial/moderately high and high degree of implementation (Grizzard et al., 2006).

3) **Scales showing that a particular programme component is implemented or not and to what degree and effort**
(Bond et al., 1997, Glanz et al., 1992, Hacker and Washington, 2004, Kansanaho et al., 2005, Lee et al., 2002, Paulsell et al., 2002, Pearson et al., 2005, Perez et al., 2011, Ryman et al., 2011, Shortell et al., 1995, Smith et al., 1977, Vinson et al., 2001).

The above categorisation between types of scales is arbitrary: studies using scale types 1 & 2 often measured programme components' implementation as well and studies using scale type 3 often converted their results into percentages.

Studies conducted the work on these implementation scales through involved site visits, personal or phone interviews with staff members of different levels (e.g., providers, managers, directors), and/or self-completion of the instruments by programme workers. Half of the studies also conducted document reviews (e.g., review of programme minutes, reports, logs, check-lists, case-records). Only about half of the studies presented psychometric characteristics of their tools (Appendix 1, Table 2. Data extraction for measurements of implementation.).

How is implementation strength score calculated, including composite implementation strength scores?

In the majority of studies mean scores were calculated for individual indicators as well as for their domains (Bonomi et al., 2002, Grizzard et al., 2006, Pearson et al., 2005, Smith et al., 1977, Teague et al., 1998, Vinson et al., 2001, Wilson et al., 2010). When overall implementation scores were presented, they were often calculated by averaging the domains' scores (Grizzard et al., 2006, Ryman et al., 2011, Shortell et al., 1995). For example Grizzard et al., have developed the hierarchical algorithm to provide overall implementation score by computing 1) averages for each response to questions within each sub-domain; 2) averages of scores of sub-domains within major domains; 3) averages of scores of major domains (Grizzard et al., 2006, Ryman et al., 2011, Shortell et al., 1995). This approach (also used by (Hebert and Veil, 2004)) "ensures that each index is weighted equally despite the fact that some indices are measured with a single question whereas others are probed with as many as five questions" (Grizzard et al., 2006).

In some cases, only average scores or percentages were provided per individual indicators and/or domains (Alkin, 1969, Bonomi et al., 2002, d'Abbs et al., 2008, Glanz et al., 1992, Gold et al., 1993, Rubin et al., 1982). In other cases, the scores were grouped or categorised to provide a descriptive level of implementation or components' implementation degree (Cooley et al., 2003). For example, some studies show whether a component is "fully or partially implemented", "not at all in place/completely in place", or whether implementation level was "low, moderate, or high" "minimal, partial, or enhanced"(Glanz et al., 1992, Grizzard et al., 2006, Hacker and Washington, 2004, Kansanaho et al., 2005, Paulsell et al., 2002, Ryman et al., 2011, Vinson et al., 2001). In cases where multiple programmes were evaluated, in some studies, programmes were subdivided

according to their overall scores. For example in the Gold et al. study, States which received top 10 overall implementation scores were ranked as States with high implementation strength, and States with low overall scores were ranked as states with low implementation strength (Gold et al., 1993).

Three studies weighted their components (Gold et al., 1993, Hebert and Veil, 2004, Ryman et al., 2011, Fixsen et al., 2005, Mowbray et al., 2003, Proctor et al., 2011, Rubin et al., 1982, Scheirer et al., 1995, Teague et al., 1998). In all of them the components and domains were weighted on the basis of their relative importance by groups of experts. If weights of domains differed, in most cases, items within domains were given equal weight. Only two studies provided detailed description why some domains were seen as more important than others (Gold et al., 1993, Hebert and Veil, 2004).

Are there examples of diverse programmes, aiming for a common outcome, where a single implementation strength score has been used?

Several studies sampled diverse programmes to test their tools. For example, Bonomi et al., scored over a hundred different organisational teams within diverse health-care systems (e.g., for profit, hospital-based, community-based) serving different populations with different diseases across the US using the ACIC tool (Bonomi, et al., 2002). The adopted ACIC tool was also used by D'Abbs et al., to score the implementation of a chronic disease strategy in primary care units serving indigenous populations in Australia (D'Abbs et al., 2008). Using the Quarterly Report Form, Orwin assessed the implementation of diverse interventions targeting different populations with substance abuse problems (Orwin, 2000). Vinson et al., have evaluated the implementation of the System-of-care model across diverse settings in the U.S. and Hawaii (Vinson, et al., 2001). Teague et al., tested their tool in fifty different programmes (organised in four groups based on their integrity to a particular service delivery model) serving people with severe mental illness (Teague et al., 1998). In all the above cases, programmes were aiming for common outcomes such as following a particular new strategy or model to reduce the burden of disease among different population groups.

Is there evidence that interventions with higher implementation strength produce better outcomes?

Only two studies provided associational information between implementation levels and study outcomes, which in many cases, was due to the specific study objectives such as developing a tool or evaluating implementation process, and not measuring implementation outcomes (Appendix 1, Table 2. Data extraction for measurements of implementation.). One study looked at the associations between implementation score and immunisation coverage within the Reach Every District Approach programme in Northern Sudan (Ryman et al., 2011), and the second looked at the level of implementation and change in physical activity among target populations (Wilson et al., 2010). Wilson et al., concluded that although outcome measures were not affected by fidelity and dose of intervention, the level of implementation was significantly associated with a change in physical activity and "practitioners should focus more energy assuring that the core components are fully implemented and be less concerned about the level of participation" (Wilson et al., 2010). The Ryman et al. study has revealed that the percentage of districts with coverage ≥80% increased as the overall RED implementation score increased. At the same time, the authors found that it was "not possible to directly attribute the overall increase in implementation coverage to RED implementation" (Ryman et al., 2011).

Summary and conclusions

Implementation process rather than outcomes

The review has shown that studies measured quantity and quality of implementation using different scaling and scoring systems, from scales with detailed definitions for each item and scale level scores to simple percentage scales. Implementation assessment was conducted using diverse methods of data collection and scoring procedures. Models, treatment protocols, guidelines, and expert opinion were all used to identify main components to be measured and their categorisation. The components often included elements of organisational structure (e.g., leadership, human resources, information systems) and processes of services delivery (activities, types, availability, and quality of services). To produce the composite implementation scores average scores for items and/or domains were calculated and presented. In some cases, the scores were further converted to percentages and/or descriptive definition of implementation degree such as: “high”, “medium” or “low level” implementation. Several studies were able to compare diverse programmes in multiple settings and geographies using the same tool. The associations between implementation level and study outcomes were presented only by two studies. In the most other cases, the research was aimed to develop and test tools or to assess the implementation process but not to look at outcomes.

Scales with detailed description provide more transparent measurements

Although the review attempted to select studies with detailed description of methodologies, some studies lacked clear definitions of the main components, how components were measured and scored, or how weights were allocated. Moreover, only half of the studies provided psychometric characteristics for their tools. Although comparing different methodologies was not the review objective, it seems that scales which included detailed description of each component and criterion for each level on the scales for specific items provided more objective and transparent measurements. These scales were also more rigorously measured in terms of their reliability and validity.

No consensus on defining and measuring implementation

Various terms and definitions were used in measuring implementation: implementation degree, implementation extent, implementation level, implementation intensity. Only one study used the term ‘implementation strength’. For ‘implementation quality,’ the terms ‘fidelity’ and ‘integrity’ were used more often. In some studies, implementation quantity was measured as a separate construct but in others it was included within the measurement of implementation fidelity. Some studies which measured fidelity of implementation also measured some or all of the following components: adherence, dosage, quality of delivery, participant responsiveness and programme differentiation. Across the studies, there was no consensus or uniform approach in defining and measuring implementation. Finally, the review has shown that there are no rigorous strategies for measuring large-scale implementation of complex interventions in low income countries.

Methodological gap— implementation strength of complex health interventions

Overall, there is a need to develop methodologies which would address the assessment of implementation strength of complex public health interventions including providing clear definitions of components, detailed tools development procedures, transparent weighting and scoring systems.

References

ALKIN, M. C. 1969. Evaluation theory development. *Evaluation comment,* 2**,** 2-7.

ARGYRIS, C., PUTNAM, R. & SMITH, D. M. 1985. *Action Science: Concepts, Methods and Skills for Research and Intervention,* San Francisco, Jossey-Bass.

BECKER, D. R. & DRAKE, R. E. 1994. Individual Placement and Support: a community mental health center approach to vocational rehabilitation. *Community Ment Health J,* 30**,** 193-206.

BECKER, D. R., SMITH, J., TANZMAN, B., DRAKE, R. E. & TREMBLAY, T. 2001. Fidelity of supported employment programs and employment outcomes. *Psychiatr Serv,* 52**,** 834-6.

BENDAVID, E., HOLMES, C. B., BHATTACHARYA, J. & MILLER, G. 2012. HIV development assistance and adult mortality in Africa. *JAMA,* 307**,** 2060-7.

BOND, G., BECKER, D., DRAKE, R. & VOGLER, K. 1997. A fidelity scale for the individual placement and support model of supported employment. *Rehabilitation Counseling Bulletin,* 40.

BOND, G. R., BECKER, D. R., DRAKE, R. E., RAPP, C. A., MEISLER, N., LEHMAN, A. F., BELL, M. D. & BLYLER, C. R. 2001. Implementing supported employment as an evidence-based practice. *Psychiatr Serv,* 52**,** 313-22.

BOND, G. R. & SALYERS, M. P. 2004. Prediction of outcome from the Dartmouth assertive community treatment fidelity scale. *CNS Spectr,* 9**,** 937-42.

BONOMI, A. E., WAGNER, E. H., GLASGOW, R. E. & VONKORFF, M. 2002. Assessment of chronic illness care (ACIC): a practical tool to measure quality improvement. *Health Serv Res,* 37**,** 791-820.

BREITENSTEIN, S. M., FOGG, L., GARVEY, C., HILL, C., RESNICK, B. & GROSS, D. 2010. Measuring implementation fidelity in a community-based parenting intervention. *Nurs Res,* 59**,** 158-65.

BRYCE, J., AMOUZOU, A., HAZEL, E., MILLER, N., JOHNS, B., GILROY, K., MUNOS, M. & VICTORA, C. 2011. Measuring the strength of implementation of community case management of childhood illness within the Catalytic Initiative to Save a Million Lives. *New working paper.* http://www.jhsph.edu/departments/international-health/IIP/projects/catalytic_initiative/wp-implementation-strength.pdf: Johns Hopkins Bloomberg School of Public Health

BUZHARDT, J., GREENWOOD, C. R., ABBOTT, M. & TAPIA, Y. 2006. Research on Scaling Up Evidence-Based Instructional Practice: Developing a Sensitive Measure of the Rate of Implementation. *Educational Technology Research and Development* 54 467-492.

CARROLL, C., PATTERSON, M., WOOD, W., BOOTH, A., RICK, J. & BALAIN, S. 2007. A conceptual framework for implementation fidelity. *Implement Sci,* 40.

CASSATA-WIDERA, A., CENTURY, J. & KIM, D. 2011. Rigorous measures of implementation: A methodological framework for evaluating innovative STEM programs

The SREE Fall 2011 Washington, D.C.: Center for Elementary Math and Science Education University of Chicago.

CHURCHMAN, D. 1979. A new approach to evaluating the implementation of innovative educational pro- grams. . *Educational Technology***,** 25-28.

COOLEY, W. C., MCALLISTER, J. W., SHERRIEB, K. & CLARK, R. E. 2003. The Medical Home Index: development and validation of a new practice-level measure of implementation of the Medical Home model. *Ambul Pediatr,* 3**,** 173-80.

COOLEY, W. C., MCALLISTER, J. W., SHERRIEB, K. & KUHLTHAU, K. 2009. Improved outcomes associated with medical home implementation in pediatric primary care. *Pediatrics,* 124**,** 358-64.

CRAMM, J. M., STRATING, M. M., TSIACHRISTAS, A. & NIEBOER, A. P. 2011. Development and validation of a short version of the Assessment of Chronic Illness Care (ACIC) in Dutch disease management programs. *Health Qual Life Outcomes,* 9**,** 49.

D'ABBS, P., SCHMIDT, B., DOUGHERTY, K. & SENIOR, K. 2008. Implementing a chronic disease strategy in two remote Indigenous Australian settings: a multi-method pilot evaluation. *Aust J Rural Health,* 16**,** 67-74.

DAMSCHRODER, L. J. & HAGEDORN, H. J. 2011. A guiding framework and approach for implementation research in substance use disorders treatment. *Psychol Addict Behav,* 25**,** 194-205.

DANE, A. V. & SCHNEIDER, B. H. 1998. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev,* 18**,** 23-45.

DURLAK, J. A. 1998a. Why program implementation is important. *In:* DURLAK J & J, F. (eds.) *Program Implementation in Preventive Trials.* Philadelphia, PA Haworth Press.

DURLAK, J. A. 1998b. Why Program Implementation Is Important? *Journal of Prevention & Intervention in the Community,* 17**,** 5-18.

DUSENBURY, L., BRANNIGAN, R., FALCO, M. & HANSEN, W. B. 2003. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Educ Res,* 18**,** 237-56.

EDWARD, A., KUMAR, B., KAKAR, F., SALEHI, A. S., BURNHAM, G. & PETERS, D. H. 2011. Configuring balanced scorecards for measuring health system performance: evidence from 5 years' evaluation in Afghanistan. *PLoS Med,* 8**,** e1001066.

EIKELAND, P. O. 2007. Downstream natural gas in Europe—High hopes dashed for upstream oil and gas companies. *Energy Policy* 35**,** 227-237.

FALLOON, I. R., ECONOMOU, M., PALLI, A., MALM, U., MIZUNO, M. & MURAKAMI, M. 2005. The clinical strategies implementation scale to measure implementation of treatment in mental health services. *Psychiatr Serv,* 56**,** 1584-90.

FIXSEN, D., NAOOM, S., BLASE, K., FRIEDMAN, R. & F., W. 2005. Implementation Research: A Synthesis of the Literature. Tampa, FL: Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.: University of South Florida.

GLANZ, K. 1985. Nutrition education for risk factor reduction and patient education: a review. *Prev Med,* 14**,** 721-52.

GLANZ, K., BREKKE, M., HARPER, D., BACHE-WIIG, M. & HUNNINGHAKE, D. B. 1992. Evaluation of implementation of a cholesterol management program in physicians' offices. *Health Educ Res,* 7**,** 151-63.

GOLD, R. B., SINGH, S. & FROST, J. 1993. The Medicaid eligibility expansions for pregnant women: evaluating the strength of state implementation efforts. *Fam Plann Perspect,* 25**,** 196-207.

GRIZZARD, T. A., BARTICK, M., NIKOLOV, M., GRIFFIN, B. A. & LEE, K. G. 2006. Policies and practices related to breastfeeding in massachusetts: hospital implementation of the ten steps to successful breastfeeding. *Matern Child Health J,* 10**,** 247-63.

HACKER, M. & WASHINGTON, M. 2004. How do we measure the implementation of large-scale change? *Measuring Business Excellence,* 8**,** 52-59.

HALL, E., ROSENTHAL, J. & WADE, J. 1994. How to make reengineering really work. *The McKinsey Quarterly,* 2**,** 107-128.

HALL, G. & LOUCKS, S. 1977. A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal,* 14**,** 263-276.

HARMS, T. & CLIFFORD, R. 1989. *Family Day Care Rating Scale,* New York, Teachers College Press.

HARMS, T., CRYER, D. & CLIFFORD, D. 1990. *Infant-Toddler Environment Rating Scale,* New York, Teachers College Press

HEBERT, R., DURAND, P. J., DUBUC, N. & TOURIGNY, A. 2003. PRISMA: a new model of integrated service delivery for the frail older people in Canada. *Int J Integr Care,* 3**,** e08.

HEBERT, R. & VEIL, A. 2004. Monitoring the degree of implementation of an integrated delivery system. *Int J Integr Care,* 4**,** 1-7.

KANSANAHO, H., PUUMALAINEN, I., VARUNKI, M., AHONEN, R. & AIRAKSINEN, M. 2005. Implementation of a professional program in Finnish community pharmacies in 2000-2002. *Patient Educ Couns,* 57**,** 272-9.

KAPLAN, R. S. & NORTON, D. P. 1992. The balanced scorecard--measures that drive performance. *Harv Bus Rev,* 70**,** 71-9.

KHAN, M. M., HOTCHKISS, D., DMYTRACZENKO, T. & ZUNAID AHSAN, K. 2012. Use of a Balanced Scorecard in strengthening health systems in developing countries: an analysis based on nationally representative Bangladesh Health Facility Survey. *Int J Health Plan Manage*.

LEE, S., CHOI, K. S., KANG, H. Y., CHO, W. & CHAE, Y. M. 2002. Assessing the factors influencing continuous quality improvement implementation: experience in Korean hospitals. *Int J Qual Health Care,* 14**,** 383-91.

LEMMENS, K. M., RUTTEN-VAN MOLKEN, M. P., CRAMM, J. M., HUIJSMAN, R., BAL, R. A. & NIEBOER, A. P. 2011. Evaluation of a large scale implementation of disease management programmes in various Dutch regions: a study protocol. *BMC Health Serv Res,* 11**,** 6.

LINNAN L & A., S. 2002. Process evaluation for public health interventions and research: an overview. *In:* LINNAN L & A., S. (eds.) *Process Evaluation for Public Health Interventions and Research.* San Francisco, CA: Jossey-Bass.

LUPI, S., VERZOLA, A., CARANDINA, G., SALANI, M., ANTONIOLI, P. & GREGORIO, P. 2011. Multidimensional evaluation of performance with experimental application of balanced scorecard: a two year experience. *Cost Eff Resour Alloc,* 9**,** 7.

MACCOLL INSTITUTE FOR HEALTHCARE INNOVATION 2000. Assessment of Chronic Illness Care Version 3.5 Seattle, WA: MacColl Institute for Health Care Innovation

MCGRAW, S., SELLERS, D., STONE, E., RESNICOW, K., KUESTER, S., FRIDINGER, F., WECHSLER, C. & WECHSLER, H. 2000. Measuring implementation of school programs and policies to promote healthy eating and physical activity among youth. *Preventive medicine,* 31**,** S86-S97.

MCGREW, J. H., BOND, G. R., DIETZEN, L. & SALYERS, M. 1994. Measuring the fidelity of implementation of a mental health program model. *J Consult Clin Psychol,* 62**,** 670-8.

MCGREW, J. H. & GRISS, M. E. 2005. Concurrent and predictive validity of two scales to assess the fidelity of implementation of supported employment. *Psychiatr Rehabil J,* 29**,** 41-7.

MCHUGO, G. J., DRAKE, R. E., WHITLEY, R., BOND, G. R., CAMPBELL, K., RAPP, C. A., GOLDMAN, H. H., LUTZ, W. J. & FINNERTY, M. T. 2007. Fidelity outcomes in the National Implementing Evidence-Based Practices Project. *Psychiatr Serv,* 58**,** 1279-84.

MEDICAL HOME INITIATIVES FOR CHILDREN WITH SPECIAL HEALTH CARE NEEDS PROJECT & ADVISORY COMMITTEE OF THE AMERICAN ACADEMY OF PEDIATRICS 2002. The Medical Home. *Pediatrics,* 109**,** 184–186

MENZIES SCHOOL OF HEALTH RESEARCH 2007. Maternal Health and Child Health Systems Assessment Tool. Version 2.1.

MOWBRAY, C. T., HOLTER, M. C., TEAGUE, G. B. & D., B. 2003. Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation,* 24**,** 315-340.

NATIONAL AIDS CONTROL ORGANIZATION, MINISTRY OF HEALTH & FAMILY WELFARE 2007. Operational Guidelines for Community Care Centres New Delhi, Delhi NCR, India.

O'BRIEN, J. L., SHORTELL, S. M., HUGHES, E. F., FOSTER, R. W., CARMAN, J. M., BOERSTLER, H. & O'CONNOR, E. J. 1995. An integrative model for organization-wide quality improvement: lessons from the field. *Qual Manag Health Care,* 3**,** 19-30.

O'DONNELL, C. 2008. Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K-12 Curriculum Intervention Research. *Review of Educational Research March* 78**,** 33-84.

ORWIN, R. G. 2000. Assessing program fidelity in substance abuse health services research. *Addiction,* 95 Suppl 3**,** S309-27.

PAULSELL, D., KISKER, E. E., LOVE, J. M. & RAIKES, H. H. 2002. Understanding Implementation in Early Head Start Programs. *Infant Mental Health Journal,* 23**,** 14-35.

PEARSON, M. L., WU, S., SCHAEFER, J., BONOMI, A. E., SHORTELL, S. M., MENDEL, P. J., MARSTELLER, J. A., LOUIS, T. A., ROSEN, M. & KEELER, E. B. 2005. Assessing the implementation of the chronic care model in quality improvement collaboratives. *Health Serv Res,* 40**,** 978-96.

PEREZ, D., LEFEVRE, P., CASTRO, M., SANCHEZ, L., TOLEDO, M. E., VANLERBERGHE, V. & VAN DER STUYFT, P. 2011. Process-oriented fidelity research assists in evaluation, adjustment and scaling-up of community-based interventions. *Health Policy Plan,* 26**,** 413-22.

PHILLIPS, S. D., BURNS, B. J., EDGAR, E. R., MUESER, K. T., LINKINS, K. W., ROSENHECK, R. A., DRAKE, R. E. & MCDONEL HERR, E. C. 2001. Moving assertive community treatment into standard practice. *Psychiatr Serv,* 52**,** 771-9.

PROCTOR, E., SILMERE, H., RAGHAVAN, R., HOVMAND, P., AARONS, G., BUNGER, A., GRIFFEY, R. & HENSLEY, M. 2011. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health,* 38**,** 65-76.

REBCHOOK, G. M., KEGELES, S. M. & HUEBNER, D. 2006. Translating research into practice: the dissemination and initial implementation of an evidence-based HIV prevention program. *AIDS Educ Prev,* 18**,** 119-36.

RIEGEL, B., LEE, C. S. & SOCHALSKI, J. 2010. Developing an Instrument to Measure Heart Failure Disease Management Program Intensity and Complexity. *Circ Cardiovasc Qual Outcomes,* 3**,** 324-330.

ROGERS, E. M. 2003. *Diffusion of innovations,* New York, Free Press.

ROSECRANS, A. M., GITTELSOHN, J., HO, L. S., HARRIS, S. B., NAQSHBANDI, M. & SHARMA, S. 2008. Process evaluation of a multi-institutional community-based program for diabetes prevention among First Nations. *Health Educ Res,* 23**,** 272-86.

ROSS, J. A. & BEGALA, J. E. 2005. Measures of strength for maternal health programs in 55 developing countries: the MNPI study. *Matern Child Health J,* 9**,** 59-70.

RUBIN, R., G, S. & REVICKI, D. 1982. A Model for Assessing the Degree of Implementation in Field-based Educational Programs. *Educational Evaluation and Policy Analysis* 4**,** 189-196.

RYMAN, T. K., ELSAYED, E. A., MUSTAFA, A. A., WIDOA, N. M., OMER, A. & KAMADJEU, R. 2011. Implementation of the reaching every district (RED) approach: experience from North Sudan. *East Mediterr Health J,* 17**,** 804-12.

SALYERS, M. P., BOND, G. R., TEAGUE, G. B., COX, J. F., SMITH, M. E., HICKS, M. L. & KOOP, J. I. 2003. Is it ACT yet? Real-world examples of evaluating the degree of implementation for assertive community treatment. *J Behav Health Serv Res,* 30**,** 304-20.

SAUNDERS, R. P., EVANS, M. H. & JOSHI, P. 2005. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. *Health Promot Pract,* 6**,** 134-47.

SCHEIRER, M. A., SHEDIAC, M. C. & CASSADY, C. E. 1995. Measuring the implementation of health promotion programs: the case of the Breast and Cervical Cancer Program in Maryland. *Health Educ Res,* 10**,** 11-25.

SHORTELL, S. M., O'BRIEN, J. L., CARMAN, J. M., FOSTER, R. W., HUGHES, E. F., BOERSTLER, H. & O'CONNOR, E. J. 1995. Assessing the impact of continuous quality improvement/total quality management: concept versus implementation. *Health Serv Res,* 30**,** 377-401.

SMITH, A. G., LOVE, J., MORRIS, M., SPENCER, L., ISPA, J. & ROSARIO, J. 1977. A Process-Evaluation of Project Developmental Continuity: findings from the PDC Implementation Study. Developmet Associates, Washington D.C.

SMITH, J. 2010. Evaluation Methods in Implementation Research: An Introduction. http://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/eis-071212.pdf: Health Service Research and Development Service.

SNETSELAAR, L. G. 1989. *Nutrition Counseling Skills: Assessment, Treatment, and Evaluation,* Rockville, MD., Aspen Publishers.

SOGARWAL, R. & BACHANI, D. 2011. Process evaluation of community care centers providing care, support, and treatment to people living with human immunodeficiency virus in India. *J Public Health Manag Pract,* 17**,** E18-23.

SOLBERG, L. I., CRAIN, A. L., SPERL-HILLEN, J. M., HROSCIKOSKI, M. C., ENGEBRETSON, K. I. & O'CONNOR, P. J. 2006. Care quality and implementation of the chronic care model: a quantitative study. *Ann Fam Med,* 4**,** 310-6.

STEURER-STEY, C., FREI, A., SCHMID-MOHLER, G., MALCOLM-KOHLER, S., ZOLLER, M. & ROSEMANN, T. 2012. The German version of the Assessment of Chronic Illness Care: instrument translation and cultural adaptation. *J Eval Clin Pract,* 18**,** 1-4.

SUMMERFELT, T. W. 2003. Program strength and fidelity in evaluation. *Applied Developmental Science,* 7**,** 55-61.

SUNAERT, P., BASTIAENS, H., FEYEN, L., SNAUWAERT, B., NOBELS, F., WENS, J., VERMEIRE, E., VAN ROYEN, P., DE MAESENEER, J., DE SUTTER, A. & WILLEMS, S. 2009. Implementation of a program for type 2 diabetes based on the Chronic Care Model in a hospital-centered health care system: "the Belgian experience". *BMC Health Serv Res,* 9**,** 152.

TALMADGE, H., RASHER, S. P., BRUN, J. K. & LEWIS, M. 1986. Partnership evaluation: collaboration between researchers and practitioners. Washington D.C.: Presented at 1986 Annual Meeting of the Society for Nutrition Education.

TEAGUE, G. B., BOND, G. R. & DRAKE, R. E. 1998. Program fidelity in assertive community treatment: development and use of a measure. *Am J Orthopsychiatry,* 68**,** 216-32.

TEAGUE, G. B., DRAKE, R. E. & ACKERSON, T. H. 1995. Evaluating use of continuous treatment teams for persons with mental illness and substance abuse. *Psychiatr Serv,* 46**,** 689-95.

TEN ASBROEK, A. H., ARAH, O. A., GEELHOED, J., CUSTERS, T., DELNOIJ, D. M. & KLAZINGA, N. S. 2004. Developing a national performance indicator framework for the Dutch health system. *Int J Qual Health Care,* 16 Suppl 1**,** i65-71.

U.S. CHAMBER OF COMMERCE 1993. The Malcolm Baldrige Award.

VICTORA, C. G., BLACK, R. E., BOERMA, J. T. & BRYCE, J. 2011. Measuring impact in the Millennium Development Goal era and beyond: a new approach to large-scale effectiveness evaluations. *Lancet,* 377**,** 85-95.

VINSON, N., BRANNAN, A. M., BAUGHMANN, L. N., WILCE, M. & GAWRON, T. 2001. The System-of-Care Model: Implementation in Twenty-seven Communities. *Journal of emotional and behavioral disorders,* 9**,** 30-42.

WAGNER, E. H., AUSTIN, B. T., DAVIS, C., HINDMARSH, M., SCHAEFER, J. & BONOMI, A. 2001. Improving chronic illness care: translating evidence into action. *Health Aff (Millwood),* 20**,** 64-78.

WAGNER, E. H., AUSTIN, B. T. & VON KORFF, M. 1996. Organizing care for patients with chronic illness. *Milbank Q,* 74**,** 511-44.

WEHMAN, P. 1988. Supported employment: Toward zero exclusion of persons with severe disabilities. *In:* S., W. P. M. M. (ed.) *Vocational rehabilitation and supported employment* Baltimore: Brooks.

WHO, WORLD BANK, GAVI & GLOBAL FUND 2010. Monitoring and evaluation of health systems strengthening: An operational framework. Geneva: WHO.

WICKIZER, T. M., WAGNER, E., CHEADLE, A., PEARSON, D., BEERY, W., MAESER, J., PSATY, B., VONKORFF, M., KOEPSELL, T., DIEHR, P. & PERRIN, E. B. 1998. Implementation of the Henry J. Kaiser Family Foundation's Community Health Promotion Grant Program: a process evaluation. *Milbank Q,* 76**,** 121-47.

WILSON, M. G., BASTA, T. B., BYNUM, B. H., DEJOY, D. M., VANDENBERG, R. J. & DISHMAN, R. K. 2010. Do intervention fidelity and dose influence outcomes? Results from the move to improve worksite physical activity program. *Health Educ Res,* 25**,** 294-305.

WORLD HEALTH ORGANIZATION 2008. Implementing the Reaching Every District Approach: A GUIDE FOR DISTRICT HEALTH MANAGEMENT TEAMS

WHO Africa Regional Office in collaboration with UNICEF and USAID's IMMUNIZATION basics Project.

WORLD HEALTH ORGANIZATION 2009. A guide to monitoring and evaluation for collaborative TB/HIV activities-2009 revision. Geneva: World Health Organization.

WORLD HEALTH ORGANIZATION & UNITED NATIONS CHILDREN'S FUND 1998. Evidence for the Ten Steps to Successful Breastfeeding. Geneva: WHO/UNICEF.

YUMO, H. A., KUABAN, C. & NEUHANN, F. 2011. WHO recommended collaborative TB/HIV activities: evaluation of implementation and performance in a rural district hospital in Cameroon. *PanAfr Med J,* 10**,** 30.

Appendix 1. Data extraction and glossary

Table 2. Data extraction for measurements of implementation.

2.1 Smith et al., 1977

Publication	First author/Year/ Title	**(Smith et al., 1977). A Process Evaluation of Project Developmental Continuity, Interim Report VII, Volume 1: Findings from the Project Developmental Continuity (PDC) Implementation Study.**
	Peer reviewed	No
Context	**Country**	USA
	Settings (urban/rural)	Head Start' programme for pre-school/elementary school age children
Background	**Aim/primary objective**	To assess the extent to which the PDC has implemented the basic PDC guidelines
	Target population	Schools involved in PDC project serving children moving from pre-school to elementary school
	Primary discipline (health, education, etc)	Education
	Research or service delivery programme	Service Delivery Programme
Conceptual basis	**Name & reference of the theoretical framework**	**PDC Program Guidelines (The Administration for Children, Youth and Families, 1974).**
Methodological development steps		IRI rating scale was developed consisting of two scale types (objective and judgemental), based on PDC Guidelines. Objective scales assessed the extent and effectiveness of implementation, this was then tested and intervals set to maximise sensitivity. Judgemental scales asked site teams to reassess programme implementation level for each component. Tool was piloted, modifications made, and the IRI finalised.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	30 subcomponents measured by 350 scales
	Nature e.g. exit interviews, checklist etc.	Interviews with PDC teachers, parents, council chairpersons and others, complimented with information from other document; site records
	Structured/semi/open-ended questionnaire	Structured and unstructured

Contents	**Thematic areas and their operational definition given in the paper**	Objective scales rated 4 areas: presence of programme activities; extent of implementation; perceived effectiveness of implementation; roles played by various groups in implementation decisions and activities. Judgemental scales measured perceived intensity/amount of effort and importance accorded to each subcomponent by programme staff. The above areas were organized across seven PDC themes and 30 subcomponents. The whole list of subcomponents and their descriptions is provided within the Smith et al., report as a separate file (pp.239-253).
Scaling	**(e.g. binary, likert scale etc)**	Likert scale/four point rating scales/categorical responses
Scoring	**(e.g. categorization of total scores of a given construct or overall tool)**	Mean objective and judgemental ratings for each IRI subcomponent and component.
Data source	**Primary**	Primary
	Secondary	Existing documents were also used to provide a comprehensive assessment of PDC implementation
Conduct	**Brief description of data collection process (by whom, timing)**	Between 1974 and 1977 structured interviews and site records, complimented with other documentation were used to complete IRI ratings from 9 sites. Site visits were also conducted by one person from Development Associates and one from High/Scope. IRI ratings were completed by the full research team.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.2 Rubin et al., 1982

Publication	First author/Year/ Title	**(Rubin et al., 1982). A model for assessing the degree of implementation in field-based educational programs**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Field-based educational programmes
Background	**Aim/primary objective**	To develop a model that measures implementation of a field-based educational programme.
	Target population	Communities implementing the Parent Education Follow Through Program
	Primary discipline (health, education, etc)	Education
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	Evaluation theory development (Alkin, 1969); A new approach to evaluating the implementation of innovative educational programs (Churchman, 1979), A developmental model for determining whether treatment is actually implemented (Hall and Loucks, 1977).
Methodological development steps		1.Major attributes of the programme identified; 2. Implementation levels specified for each implementation objective; 3. Identification made whether criterion level achieved.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	Dependent on educational programme (none described for the given example)
	Nature e.g. exit interviews, checklist etc.	Dependent on programme activity/attribute measuring i.e. observations attendance sheets, sign-in sheets, interviews.
	Structured/semi/open-ended questionnaire	Structured
Contents	**Thematic areas and their operational definition given in the paper**	Activities and attributes associated with the programme (e.g. home visits of program children by paraprofessionals; time paraprofessionals spent in institutional activities; parents attendance at committees meetings)
Scaling	**(e.g. binary, likert scale etc)**	Scale (%)

Scoring	**(e.g. categorization of total scores of a given construct or overall tool)**	% of each programme attribute measured against criterion level. Scored as yes/no based on whether the criterion % is met.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	The rating scale was completed at two levels by programme staff members and by external evaluators (who visit school twice in a school year) and results compared.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.3 Glanz et al., 1992

Publication	First author/Year/ Title	**(Glanz et al., 1992). Evaluation of implementation of a cholesterol management program in physicians' offices.**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Family physician office practices implementing cholesterol management programmes
Background	**Aim/primary objective**	To describe an evaluation of the implementation of a pilot programme for cholesterol management in family physicians' offices.
	Target population	Primary care clinics targeting people with high cholesterol levels
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	Partnership evaluation: collaboration between researchers and practitioners(Talmadge et al., 1986). Action science. Concepts, Methods and Skills for Research and Intervention (Argyris et al., 1985).
Methodological development steps		1. Operational measures of variables were defined 3. A scale was developed to score programme implementation 4. Qualitative data helped establish the validity of responses to quantitative items
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	14
	Nature e.g. exit interviews, checklist etc.	Clinic programme descriptions, clinic staff interviews, patient care data, clinic contact records and archival data.
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1) Screening: measurement of blood cholesterol, Identification of other cardiovascular risk factors, assignment to receive treatment if warranted, physician message; 2)counselling: assessment of eating pattern, informing patient of cholesterol value and risk level, discussion of eating pattern assessment results, instruction on cholesterol-lowering diet, discussion of and setting specific eating behaviour goals; monitoring cholesterol and eating pattern changes 3)tracking and monitoring: tracking patient appointments and follow-up visits, tracking billing and reimbursement for patient counselling, recording and summarising CMP progress,

		creating communication networks to keep physicians and office staff informed of CMP progress
Scaling	**(e.g. binary, likert scale etc)**	Scale 1-5 for programme implementation
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	1 = 'not at all in place', 5 = 'completely in place' for programme implementation scale
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Each clinic provided a description of its cholesterol management programmes within 2 months after program start-up and after 6 months. Clinic staff semi-structured interviews were conducted. Telephone interviews were conducted in February 1988. Copies of patient visits were collected during the 6 month period. Additional archival data were collected.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.4 Gold et al., 1993

Publication	First author/Year/ Title	**(Gold et al., 1993). The Medicaid eligibility expansions for pregnant women: evaluating the strength of state implementation efforts.**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	State Medicaid agencies and maternal and child health programs
Background	**Aim/primary objective**	To measure the aggressiveness with which states have implemented the expansions of Medicaid eligibility; to examine factors that might explain why some states have been more aggressive than others
	Target population	State programmes serving pregnant women in need
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	None
Methodological development steps		1. Two-part survey developed; 2. additional data to act as a baseline gathered; 3. measures undertaken by states summarised into indices reflecting four main aspects of the Medicaid expansions; 4. these indices combined into single index of overall strength of implementation (ISI); 5. weights assigned for each thematic area; 6. items within each index grouped according to two or three major areas to form components within each index.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	53
	Nature e.g. exit interviews, checklist etc.	Two surveys, telephone interviews, records of Medicaid, additional data (e.g. the proportion of low-birth-weight babies, the proportion of women of reproductive age below 185% of poverty)
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1. Eligibility and enrolment: changing Medicaid income eligibility levels, easing the enrolment process by employing at least one strategy to facilitate enrolment, offering presumptive eligibility, out-stationing enrolment to health care sites; 2.Outreach activities: informing pregnant low-income women that they might be

		eligible for subsidised care through variety of sources (e.g. media, posting, hotlines); 3.Enhanced prenatal care service: expanding the benefits package for pregnant women by offering them financial coverage; 4. Reimbursement generosity: increasing the amount that state Medicaid programs pay health care providers for prenatal care and delivery, increasing either the number of providers willing to accept Medicaid patients or the ability of public providers to expand their services.
Scaling	**(e.g. binary, likert scale etc)**	5 points scale to rank the states; 0-100 scale for the overall ISI
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	"States were ranked according to their score on each of the four indices, and on the overall ISI. States were divided into five groups, with the top-ranking 10 states receiving a rank of 1 and the lowest-ranking 10 receiving a rank of 5...For the overall index, a score of 0 indicated the least overall effort and a score of 100 indicated the most possible effort."
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Surveys mailed to directors of state Medicaid agencies and MCH programme directors with follow-up mailing to states that did not respond within one month. Between January and April 1992, non-respondents were contacted by telephone, as were state directors whose responses lacked data on key questions. The study was conducted in 1991-1992
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.5 Shortell et al., 1995

Publication	First author/Year/ Title	**(Shortell et al., 1995). Assessing the Impact of Continuous Quality Improvements/Total Quality Management: Concept versus Implementation**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	U.S. hospitals implementing continuous quality improvement and total quality management programmes (CQI/TQM)
Background	**Aim/primary objective**	To examine the degree of quality improvement (QI) implementation and investigate factors influencing the implementation of QI activities
	Target population	U.S. hospitals
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service Delivery Programme
Conceptual basis	**Name & reference of the theoretical framework**	Baldrige Award criteria (U.S. Chamber of Commerce, 1993).
Methodological development steps		QI implementation scales were developed based on the Baldridge Award criteria.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	The Cronbach alpha reliabilities were: leadership (a = .93); information and analysis (a = .86); human resources utilisation-empowerment (a = .80); human resources utilisation-education and training (a = .79); strategic quality planning (a = .88) and quality management (a = .85).
Structure	**No. of items**	Organisational culture = 20; implementation approach = 6 areas
	Nature e.g. exit interviews, checklist etc.	Organisational culture – self-administered questionnaire to hospital employees. Implementation approach - questionnaire completed by senior executives, QI council members and quality assurance committee members
	Structured/semi/open-ended questionnaire	Not specified
Contents	**Thematic areas and their operational definition given in the paper**	Degree of implementation is measured as a function of hospital size, culture, implementation approach and whether or not the hospital is involved in CQI/TCM. CQI/TCm principles are: leadership; information and analysis; human resources utilization - empowerment; human resources utilization - education and training; strategic quality planning; quality management.

Scaling	**(e.g. binary, likert scale etc)**	Implementation approach questionnaire and performance measures used ordinal scales (1-7).
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Means and standard deviations were calculated for each variable i.e. Scales.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Organisational culture, implementation approach and QI implementation were measured using administered questionnaires. Performance measures was assessed by hospital CEOs and directors and objective measures of clinical efficiency were collected independently.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.6 Bond et al., 1997

Publication	First author/Year/ Title	**(Bond et al., 1997). A fidelity scale for the individual placement and support model of supported employment (SE).**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Programs serving the people with severe mental illness in three USA regions
Background	**Aim/primary objective**	To develop and test the fidelity scale for SE model
	Target population	Programs serving the people with severe mental illness helping people achieve paid employment
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	The Individual Placement and Support (IPS); (Becker and Drake, 1994)of supported employment model (Wehman, 1988).
Methodological development steps		1. The IPS Fidelity Scale was developed through a semi-structured interview with knowledgeable staff. The IPS manual, brainstorming sessions among the authors, and the SE literature were used to create a content of its' items. 2.To test the scale three types of programs were selected: IPS programs, other types of SE programs, and non-SE vocational rehabilitation programs
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Interrater reliability was 0.80 or higher. Internal consistency reliability (Cronbach's alpha) for the consensus ratings for total scale was 0.92, staffing - 0.72, organization-0.65, service-0.90. Construct validity for the total scale was significant (F, p=0.001)
Structure	**No. of items**	15
	Nature e.g. exit interviews, checklist etc.	Semi-structured interviews with a knowledgably staff worker, as well as agency records and direct observation
	Structured/semi/open-ended questionnaire	Semi-structured
Contents	**Thematic areas and their operational definition given in the paper**	Staffing (caseload size and types of services provided); Organization (contact with mental health team; working as a unit, types of supervision); and Services (continuous assessment, types of jobs provided, community support, outreach). The content of thematic areas is provided within an instrument in Appendix 1.
Scaling	**(e.g. binary, likert scale etc)**	Scale 1-5

Scoring	**(e.g. categorization of total scores of a given construct or overall tool)**	All item values were summed with a range of values as follows: Total scale (15-75), Staffing (3-15), Organization (3-15), and service (9-45). The Total scale was used to yield a categorical fidelity classification (>65 = consistent with IPS, 56-65 = partially consistent with IPS, <56 = not IPS)
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Interviews (one hour each) with staff workers were conducted by six researchers in 27 sites. Program supervisors were interviewed via in-person or telephone interview.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None in this article. But in a Bond et al., 2001 review of use of the IPS Scale, nine out of ten studies found positive associations with employment outcomes (Bond et al., 2001).
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	The use of the tool is summarised by Bond et al., 2001 Measurement of Fidelity of Implementation of Evidence-Based Practices: Case Example of the IPS Fidelity Scale (Bond et al., 2001). E.g., (Becker et al., 2001, McGrew and Griss, 2005)

2.7 Teague et al., 1998

Publication	First author/Year/ Title	**(Teague et al., 1998). Program fidelity in assertive community treatment: development and use of a measure.**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	A complex community-based service, helping people with severe mental disorders
Background	**Aim/primary objective**	To describe the development and results of usage of the Dartmouth ACT Scale (DACTS) as a measure of program fidelity to ACT
	Target population	programmes serving people with severe mental illness
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	Evaluating use of continuous treatment teams for persons with mental illness or substance abuse (Teague et al., 1995).
Methodological development steps		1. Program criteria were developed 2. Criteria were grouped into three dimensions: human resources structure/composition, organisational boundaries, and nature of services 3.For each criteria indicators were defined 4. Anchors for each indicator were specified on the rating scale.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Cronbach's alpha for the 33 programs reporting all items was 0.92.
Structure	**No. of items**	26
	Nature e.g. exit interviews, checklist etc.	Staff reports on program behaviour, documents review including quantitative data on staffing, clientele, and services, structured interviews.
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1. Human resources –"addresses composition and structure of program staffing" (e.g. ration of clients to staff). 2.Organisational boundaries – "addresses programme responsibility and relationships with other programmatic components". 3.Nature of services-"addresses the range and nature of services and overall treatment approach"
Scaling	**(e.g. binary, likert scale etc)**	Scale 1-5

Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Scoring description was done for each item separately from "low"=1 to "high"=5. For example, "the low anchor for annual caseload retention was defined as 50%."
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Fifty programs were rated between 1995 and 1996. Final ratings were done by the central research team.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	Instrument widely used in substance abuse, psychiatric rehabilitation and mental health services (Bond and Salyers, 2004, McHugo et al., 2007, Phillips et al., 2001, Salyers et al., 2003).

2.8 Orwin et al., 2000

Publication	First author/Year/ Title	**(Orwin, 2000). Methodological challenges in study design and implementation**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Multisite substance abuse rehabilitation projects for homeless persons
Background	**Aim/primary objective**	To explore assessment of treatment fidelity in substance abuse services research
	Target population	Programmes serving homeless substance abusers
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	A variety of frameworks related to implementation research used (e.g. logic models, formative studies)
Methodological development steps		1. Quarterly Report Form (QRF) developed to collect data at programme- and participant-level; 2. 39 distinct services identified and glossary of service activities developed; 3. services classified by setting, provider type and funding source; 4. implementation histories developed for each project by (i) establishing key events and (ii) using QRF; 5. planned services map created; 6. implementation, fidelity and leakage scales calculated.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	List of service activities agreed by expert panel; key event chronology agreed by senior staff member.
Structure	**No. of items**	39 items. Some items are tools (e.g. outpatient visit concentration index)
	Nature e.g. exit interviews, checklist etc.	Semi-structured interviews with knowledgeable staff worker, agency records, direct observation, bimonthly phone calls, annual site visits
	Structured/semi/open-ended questionnaire	Semi-structured
Contents	**Thematic areas and their operational definition given in the paper**	1) Objectives: (a) were facilities acquired as intended?(b) were target sample sizes achieved? (c) were specific subgroups recruited as intended? (d) were planned service components successfully implemented? (e) was the intensity or “dose” of the intervention consistent with the program model? (f) was the intervention stable once implemented? and (g) were systems-level objectives met?; 2)barriers to achieving these objectives; 3) efforts undertaken to remove or reduce barriers. The areas also covered

Scaling	**(e.g. binary, likert scale etc)**	1)The total-services scale was developed from the QRF services data to provide a rough summary indicator of overall intervention strength at the participant level. It represents the number of services received multiplied by the duration of the service. 2) The fidelity scale was computed like the total services scale, but is restricted to planned services. 3) the leakage scale captured the degree to which participants were provided services planned only for other intervention groups. The formulas for the calculation of all scale are provided in Appendix 3, **Error! Reference source not ound.**.
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Fidelity scale: All item values were summed with a range of values ranging from 3 - 45 points. Total scale was used to yield a categorical fidelity classification presented as a percentage (>65 = consistent with IPS, 56-65 = partially consistent with IPS, <56 = not IPS)
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	1. QRF filled in by either the study authors or the project coordinator - quarterly. 2.National evaluation team was examining the logic models. The project was conducted between September 1990 and 1993
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.9 Vinson et al., 2001.

Publication	First author/Year/ Title	**(Vinson et al., 2001). The system-of-care model: Implementation in twenty-seven communities**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Community based children's mental health services
Background	**Aim/primary objective**	To evaluate implementation of the system-of-care model - the Comprehensive Community Mental Health Services for Children and their Families Program.
	Target population	Communities receiving a grant as part of the the Comprehensive Community Mental Health Services for Children and their Families Program.
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service Delivery Programme
Conceptual basis	**Name & reference of the theoretical framework**	Own framework was developed based on a review of the literature to identify the key features and functions of mental health services, service system management, coordination of services.
Methodological development steps		1.A framework, data collection tools, and an evaluation protocol were developed and pilot-tested. 2. Attributes were devised to capture key characteristics (see thematic areas) and each attribute was divided into key elements specific to that attribute. 3. Index developed for the framework to quantitatively rate system attributes via a series of questions (see scoring).
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Internal consistency high (α ≤0.94); Cronbach's alpha coefficients for 12/16 attributes > 0.69 but inadequate internal consistency for target population, human resources development, evaluation and funding.
Structure	**No. of items**	16 attributes; 100 questions
	Nature e.g. exit interviews, checklist etc.	Interviews, observational notes, sites documentation review
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	Two major domains: 1)Infrastructure (service array; goals and vision; case review; community based; interagency structure; communication; evaluation; target population; human resources development; funding) and 2: Service delivery (cultural competence; case management; family focus; service delivery coordination; individualised services plan; services accessibility).
Scaling	**(e.g. binary, likert scale etc)**	Scale 0-3.

Scoring	**(e.g. categorization of total scores of a given construct or overall tool)**	Scales were scored as follows: 0 = absence/minimal presence of that item; 3 = mature and systematic presence. Scale items for each attribute were averaged for a total attribute score. A total score for each domain and a composite system-of-care development score were calculated.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Annual visits to 27 sites were made between 1995 and 1998. Qualitative data was derived from multiple sources. After review of site documents and summary tables, two trained site visitors completed a 3 to 4 day site visit to collect data via interviews and review case records. After each site visit, findings were complied into a comprehensive report.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.10 Bonomi et al., 2002

Publication	First author/Year/ Title	**(Bonomi et al., 2002). Assessment of Chronic Illness Care (ACIC): A practical Tool to Measure Quality Improvement**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Organisations/health-care systems delivering care to patients with chronic illnesses
Background	**Aim/primary objective**	To describe the development and testing of the ACIC in health care systems
	Target population	Healthcare facilities treating/caring for chronic disease patients
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	The Chronic Care Model (Wagner et al., 2001, Wagner et al., 1996).
Methodological development steps		1. The ACIC was developed based on six areas of system change suggested by CCM. 2. To test the ACIC organisational teams implemented interventions in each of the six areas of the CCM and completed the ACIC. 3.The process was assessed independently by two faculty members
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Paired t-tests measured sensitivity of the ACIC - all 6 subscale scores were responsive to system improvement (P<0.05). Pearsons correlations with the ACIC subscales and faculty ratings ranged from 0.28 to 0.52
Structure	**No. of items**	28
	Nature e.g. exit interviews, checklist etc.	Pre-post, self-report data from organisational teams; group meetings, independent faculty rating
	Structured/semi/open-ended questionnaire	Structured
Contents	**Thematic areas and their operational definition given in the paper**	1)Linkages to community resources: linking patients to outside resources, partnerships with community organisations, linkages to regional health plans; 2)self-management support: assessment and documentation of self-management needs and activities; addressing concerns of patients and families, effective behaviour change interventions and peer support; 3)decision support: evidence-based guidelines , involvement of specialists in improving primary care , provider education for chronic Illness care, informing patients about guidelines; 4) delivery system design: addressing practice team functioning and leadership, appointment system, follow-up, continuity of care, planned visits for chronic illness care 5) clinical

		information systems: availability and a content of registry, reminders to providers and a feedback, information about relevant subgroups of patients needing services and patient treatment plans 6)organisation of the health system: leadership, improvement strategies, incentives and regulations. The content of items in thematic areas is provided within the article (pages 811-820).
Scaling	**(e.g. binary, likert scale etc.)**	Scale 0-11
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Response to each of the 28 items fall within four descriptive levels of implementation ('little or none' to 'fully implemented intervention'). Within each of the four levels, respondents choose one of three ratings based on a 0-11 subscale: 0–2 (little or no support for chronic illness care); 3–5 (basic or intermediate support for chronic illness care); 6–8 (advanced support); and 9–11 (optimal, or comprehensive, integrated care for chronic illness). Subscale scores for the 6 areas are derived by summing responses for items in that subsection and dividing by the corresponding number of items.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	108 organisational teams completed the ACIC (15-30 minutes) at the beginning and end of the collaborative (after 13 months) using a format that produced average ratings of their system's approach to delivering care for the targeted chronic condition. To test the ACIC two members independently assessed the progress of each team using a single five-point rating based on cumulative reports prepared by teams.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	The tool is widely used in chronic care of variety of chronic illnesses mainly in the USA and Europe (Cramm et al., 2011, Lemmens et al., 2011, Pearson et al., 2005, Solberg et al., 2006, Steurer-Stey et al., 2012, Sunaert et al., 2009)

2.11 Lee et al., 2002

Publication	First author/Year/ Title	**(Lee et al., 2002). Assessing the factors influencing continuous quality improvement (CQI) implementation: experience in Korean hospitals.**
	Peer reviewed	Yes
Context	**Country**	Korea
	Settings (urban/rural)	Public and teaching hospitals
Background	**Aim/primary objective**	To assess the extent of continuous quality improvement (CQI) implementation
	Target population	Institutions developing strategies to improve CQI implementation
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	Concept of continuous quality improvement pyramid. An integratIve model for organisation-wide quality improvement (O'Brien et al., 1995). Malcolm Baldrige National Quality Award Criteria (MacColl Institute for Healthcare Innovation, 2000, Shortell et al., 1995).
Methodological development steps		1. To measure implementation degree Malcolm Baldrige National Quality Award Criteria was used ; 2. extent to which the organisation fulfilled each of four components of the CQI pyramid was assessed; 3. instrument pilot-tested among the directors from quality improvement departments.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Cronbach's alpha test for internal consistency and reliability was 0.75-0.93
Structure	**No. of items**	7 for the implementation section of the questionnaire
	Nature e.g. exit interviews, checklist etc.	Interviews, self-completed questionnaire
	Structured/semi/open-ended questionnaire	Structured
Contents	**Thematic areas and their operational definition given in the paper**	Thematic areas for Implementation assessment was derived from Malcolm Baldrige Quality Award Criteria: 1.Leadership; 2.Strategic quality planning; 3.Customer satisfaction; 4.Infromation and analysis; 5.Human resources management; 6.Quality management; 7.Organisational performance results. The thematic areas are not defined.
Scaling	**(e.g. binary, likert scale etc)**	5 point scale

Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	1=strong disagreement, 2=disagreement, 3=neutral, 4=agreement, 5=strong agreement. Mean scores across seven dimension were calculated
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	The survey was completed by the directors of CQI departments between September and October 2000
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.12 Pausell et al., 2002

Publication	First author/Year/ Title	**(Paulsell et al., 2002). Understanding implementation in Early Head Start programmes: Implications for policy and practice**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Early Head Start' programme: child development services
Background	**Aim/primary objective**	To assess the extent of programme implementation and the quality of key child development services in 17 research programmes
	Target population	Services involved in Early Head Start programmes
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	Authors relied on two measures: Toddler Environment Rating Scale (Harms et al., 1990), and Family Day Care Rating Scale (Harms and Clifford, 1989).
Methodological development steps		1. A set of detailed rating scales and a process for rating programmes' implementation of key elements was developed. 2. A large amount of qualitative information about program implementation was synthesised into a concise set of ratings for each program.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Rating scale (based on programme requirements) reviewed by representatives from the Head Start Bureau, Early Head Start technical assistance network and Early Head Start Research Consortium
Structure	**No. of items**	3 main areas with 25 programme elements
	Nature e.g. exit interviews, checklist etc.	Interviews, self-completed questionnaire
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1) Early childhood development and health services: frequency of services, developmental assessments, follow-up services, child health services, parent involvement, individualisation of services, group socialisation activities; 2) family and community partnerships: individualised family partnership agreements, availability of services, frequency of family development services, parent involvement, collaborative relationships with other service providers, advisory committees, transition plans; 3) Management systems and procedures: staff development and

		programme management : supervision, training, staff retention, compensation, staff morale, policy council, communication systems, goals, objectives, and plans, self-assessment, community needs assessment
Scaling	**(e.g. binary, likert scale etc)**	5-point scale for each programme element
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	1 = minimal implementation; 2=low-level implementation; 3=moderate implementation; 4= full implementation; 5 = enhanced implementation
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	National evaluation team and 15 local research teams conducted the evaluation during two rounds of site visits between 1997 and 1999. Evaluators conducted individual and group interviews, reviewed case files and programme documents, distributed and collected self-administered surveys, and observed service delivery.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.13 Cooley et al., 2003

Publication	First author/Year/ Title	**(Cooley et al., 2003). The Medical Home Index: Development and Validation of a New Practice-level Measure of Implementation of the Medical Home Model**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Paediatric primary healthcare services for children with special healthcare needs (CSHCN)
Background	**Aim/primary objective**	To develop and validate a quality improvement tool to measure the Medical Home health care model.
	Target population	Clinical practices serving children with special health care needs in seven states
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	A Medical home framework (Medical Home Initiatives for Children With Special Health Care Needs Project and Advisory Committee of the American Academy of Pediatrics, 2002). Assessment of Chronic Illness Care (ACIC): A Practical Tool to Measure Quality Improvement.(Bonomi et al., 2002).
Methodological development steps		1. National experts agreed on the 6 MHI domains and determined the themes within the domains. 2. 25 themes were divided into 6 domains of practice activity critical to the quality of care in a medical home. 3. Project staff visited the office environment and reviewed specific care elements to provide the basis for the MHI scores. 4. The instrument was tested on a national sample of pediatric practices following the same procedure.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Interrater reliability: Kappa coefficients were above .50 for all 25 themes. Intraclass correlation coefficients were 0.98 (between interviewers) and 0.97 (between interviewer and practice scores). Internal consistency reliability standardised alpha coefficients across the 6 domains of the MHI ranged from .81 to 0.91, and the overall standardised alpha coefficient was .96.
Structure	**No. of items**	25
	Nature e.g. exit interviews, checklist etc.	Interviews, self-completed questionnaire
	Structured/semi/open-ended questionnaire	Structured

Contents	**Thematic areas and their operational definition given in the paper**	1) Organisational capacity: mission of the practice, communication/access, access to medical records, office environment, family feedback, cultural competence, staff education; 2)chronic condition management: identification of CSHCN, care continuity, continuity across settings, cooperative management with specialists, supporting transition to adult services, family support; 3)care coordination: role definition, family involvement, child and family education, assessment of needs/plans of care, resource information and referrals, advocacy; 4) community outreach: community assessment of needs of CSHCN, community outreach to agencies and schools; 5)data management: electronic data support, data retrieval capacity; and 6) quality improvement: quality standards (structures), quality activities (processes).
Scaling	(e.g. binary, likert scale etc)	Scale 0-100
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Each theme is scored across 4 levels of achievement. The 4 levels of achievement for each theme correspond to a continuum of quality starting with basic pediatric care (level 1) through responsive care (level 2), proactive care (level 3), and comprehensive care (level 4). Each theme can receive a score from 1 (partial achievement of level 1) to 8 (complete achievement of level 4). The raw summary scores with a range of 25–200 were transformed to a scale of 0–100 to facilitate interpretation of scores.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Two researchers conducted on-site 90 minutes interviews with primary care officers, pediatric care staff completed the instrument. The study was conducted between 2000 and 2001.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	Not reported in this study. However, in Cooley, 2009 Higher MHI scores were associated with significantly fewer hospitalisations (Cooley et al., 2009).
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	(Cooley et al., 2009)

2.14 Hacker & Washington, 2004

Publication	First author/Year/ Title	**(Hacker and Washington, 2004). Emerald Article: How do we measure the implementation of large-scale change?**
	Peer reviewed	Yes
Context	**Country**	Botswana
	Settings (urban/rural)	Public service organisational change projects
Background	**Aim/primary objective**	To create a tool to assess implementation of a large-scale performance management system
	Target population	Public service organisations implementing large-scale performance management system
	Primary discipline (health, education, etc)	Public/Civil service
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	None
Methodological development steps		1. A review on implementation literature conducted to define the main areas that can be evaluated the success of a large-scale implementation effort; 2. Main six areas defined and operationalised; 3. questionnaire developed to be used in evaluation of the implementation of any large-scale organisational change project
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	6 themes, 42 questions
	Nature e.g. exit interviews, checklist etc.	Survey
	Structured/semi/open-ended questionnaire	Structured

Contents	**Thematic areas and their operational definition given in the paper**	1. Well defined key result areas and goals: the goals at the unit level should be linked to the overall vision of the organisation, the goals are appropriate and measurable; 2. Well defined objectives: using appropriate, well-defined strategies to achieve the established goals, linking strategies to the overall vision, translating strategic goals to annual performance objectives, deploying the objectives down through the organisation, evaluating objectives and strategies; 3. Well defined measurement process: evaluation of the management system to insure that correct measures are in place, staff understanding of how to create and interpret the data; 4. Well established reviews: consistently review systems to determine whether activities are achieving required results and to provide a feedback of the decision-making process; 5. Responsibilities defined and visible and 6. Continuous improvement processes exist: documenting, training and monitoring results to insure the standardisation of the improvement effort, understanding who are the actors and what their roles in the change process.
Scaling	**(e.g. binary, likert scale etc)**	Scale 1-7
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	1=Implementation has not started in organisation/7=Implementation complete across Ministry. Average of the six items in the survey was calculated.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	22 change agents from the 16 ministries in the government of Botswana were interviewed by the authors, documentation was reviewed and large-scale assessment survey was completed.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.15 Hebert et al., 2004

Publication	First author/Year/ Title	**(Hebert and Veil, 2004). Monitoring the degree of implementation of an integrated delivery system (IDS)**
	Peer reviewed	Yes
Context	**Country**	Canada
	Settings (urban/rural)	Integrated service delivery programme for frail elderly (rural & urban)
Background	**Aim/primary objective**	To develop a methodology to rate the degree of implementation and to monitor the implementation of an IDS system
	Target population	Services caring for frail elderly
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	PRISMA model (Hebert et al., 2003).
Methodological development steps		1. Measurable indicators generated for each component and approved by two committees; 2. data collected every six months to assess the indicators; 3. indicators rated according to two sources of data; 4. single entry point indicators completed using statistics from the local agencies where they were based; 5. relative importance of each indicator was weighted
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Indicators discussed and approved by two committees of experts
Structure	**No. of items**	20
	Nature e.g. exit interviews, checklist etc.	Interviews, focus groups, surveys, documentation and clinical data analysis, observations, management data monitoring
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1. Coordination of all organisations involved in delivering health and social services: coordination must be established at the strategic, tactical and operational levels of the organisations; 2. a single entry point: the mechanism for accessing the services of all the health care institutions and community organisations in the area for the frail senior with complex needs.; 3. case management: variation between actual number of case managers and number needed according to proportion of senior citizens in the area, variation between actual average caseload and recommended caseload; 4. a single assessment tool with a case-mix classification system: evaluation of the

		needs of clients in all organisations and by all the professionals working either in home care organisations or in hospitals and institutions; 5. an individualised service plan: percent utilization of the individualised service plan by case managers 6. a computerised clinical chart: availability of a computer program for sharing clinical information in real time, sufficient number of computers for all partners, utilization of the computerised computer chart by partners
Scaling	**(e.g. binary, likert scale etc)**	None
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Each indicator given a maximum score of between 3 - 10 points. Points were added, to a maximum of 100. Implementation degree was provided in percentages
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Data from policy-makers, managers, clinicians, clients and informal care-givers collected every 6 months over 30 months in three experimental areas by a team of researchers from July 2001 - July 2003
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.16 Kansaharo et al., 2005

Publication	First author/Year/ Title	**(Kansanaho et al., 2005). Implementation of a professional program in Finnish community pharmacies in 2000-2002.**
	Peer reviewed	Yes
Context	**Country**	Finland
	Settings (urban/rural)	Community-based pharmacies
Background	**Aim/primary objective**	To assess implementation of a national project (TIPPA) to promote patient counselling in community pharmacies
	Target population	Staff from registered community pharmacies registered in TIPPA
	Primary discipline (health, education, etc)	Pharmacy
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	None
Methodological development steps		Questionnaire developed to measure pharmacies' implementation rate in patient counselling
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	16 questions and 2 open-ended questions
	Nature e.g. exit interviews, checklist etc.	Self-completed (postal) questionnaire
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	One major area, 'patient counselling' by pharmacies was evaluated. The authors do not provide an operational definition. The analysis of questionnaire items provided following areas: staff training and involvement, information sources, privacy for patients, availability of contracts and action plans, linkages to local health care providers, and feedback from the customers.
Scaling	**(e.g. binary, likert scale etc)**	5 point scale
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Sum scales of the implementation rate calculated and scored "low" (0-15 points), "moderate" (16-31 points) and "high implementation rate" (32-48 points).
Data source	**Primary**	Primary

	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	734 registered pharmacists completed a postal questionnaires in June 2002.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.17 Pearson et al., 2005

Publication	First author/Year/ Title	**(Pearson et al., 2005). Assessing the implementation of the chronic care model (CCM) in quality improvement collaborative.**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Clinical chronic illness care (2 nationwide collaboratives; 1 regional collaborative)
Background	**Aim/primary objective**	To measure organisations' implementation of CCM interventions for chronic care quality improvement (QI)
	Target population	42 health care organisations serving people with chronic illness
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	Authors suggested their own framework in evaluating implementation of "the change activities". They created a catergorisation scheme which formed "the basis for assessing the alignment of intervention change activities with six CCM elements" (see in #15).
Methodological development steps		*only implementation intensity measures are reported: 1.Two indicators of CCM implementation intensity created: (i) total count of the organisational change activities; (ii) qualitative rating of the depth of the change; 2. three levels of depth of change activities in systems or practices created; 3. depth ratings presented as percents of the maximum rating in each category; 4. sites completed own assessments on integrity to CCM.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Counts and depth ratings were significantly correlated overall (r=0.80, p. <001) and at each of the six element levels (r=0.67 to 0.70; p.<0001). Intercoder reliability at the initial assessment showed 82% and 100% in a final agreement. Correlation between implementation depth ratings and ACIC scores was as follows: in general, depth of change efforts negatively correlated with site's self-assessment prior to the intervention (ACIC baseline scores) and positively correlated after the intervention (ACIC follow-up scores.
Structure	**No. of items**	23
	Nature e.g. exit interviews, checklist etc.	Organisation documents, final reports, exit interviews
	Structured/semi/open-ended questionnaire	Structured and unstructured

Contents	**Thematic areas and their operational definition given in the paper**	1.Delivery system redesign: care management roles, team practice, care delivery/coordination, visits and follow-up; 2.self-management support: patient education and support, self-management assessment, resources and tools, decision making, guidelines available for patients; 3.decision support: guideline institutionalisation and prompts, provider education, expert consultation and support; 4.information systems: patient registry system, use of information for care management, feedback of performance data; 5.community linkages for patients and for community; 6.health system support: leadership support, provider participation, coherent system improvement, spread.
Scaling	**(e.g. binary, likert scale etc)**	3 point scale for implementation intensity
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	2=change activity likely to have an impact; 0=no impact. Overall depth ratings for each site presented as % of the max rating (46 points) possible in each category
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Reviews of organisational documents, interviews with team leaders. The reports and documents were coded by two independent researchers. All data collection was conducted in 1999-2002.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.18 Grizzard et al., 2006

Publication	First author/Year/ Title	**(Grizzard et al., 2006). Policies and practices related to breastfeeding in Massachusetts: hospital implementation of the ten steps to successful breastfeeding**
	Peer reviewed	Yes
Context	**Country**	USA
	Settings (urban/rural)	Rural and urban maternity hospitals in Massachusetts, USA
Background	**Aim/primary objective**	To assess implementation and related factors of WHO/UNICEF 10 steps to successful breastfeeding
	Target population	Hospitals providing maternity care and implementing the 10 steps
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	Tool based on the WHO 10 steps for breastfeeding(World Health Organization and United Nations Children's Fund, 1998).
Methodological development steps		1. A c/s survey was designed and survey questions generated using a list of indices of each step; 2. A series of indices for each of the 10 steps were selected using literature regarding important differences between hospitals fully and non-fully implementing the steps; 3. Indices evaluated for relevance and completeness by a 25-person committed of health professionals
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Indices checked and assessed by a 25-person committee of experts. Study questions piloted in 1 hospital and revised accordingly.
Structure	**No. of items**	55
	Nature e.g. exit interviews, checklist etc.	Phone survey
	Structured/semi/open-ended questionnaire	Structured

Contents	**Thematic areas and their operational definition given in the paper**	Thematic areas reflected the 10 UNICEFF steps: 1. Have a written breastfeeding policy that is routinely communicated to all healthcare staff 2. Train all healthcare staff in skills necessary to implement this policy 3. Inform all pregnant women about the benefits and management of breastfeeding 4. Help mothers initiate breastfeeding within a half-hour of Birth 5. Show mothers how to breastfeed, and how to maintain lactation even if they should be separated from their infants 6. Give newborn infants no food or drink other than breast milk, unless medically indicated 7. Practice rooming-in—allow mothers and infants to remain together—24 h a day 8. Encourage breastfeeding on demand 9. Give no artificial teats or pacifiers to breastfeeding infants 10. Foster the establishment of breastfeeding support groups and refer mothers to them on discharge from the hospital or clinic.
Scaling	**(e.g. binary, likert scale etc)**	Scale 1-5; yes/no
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Algorithm developed to quantify and score implementation. Each response scored out of 100%. Step and overall implementation scores computed by averaging over the response scores for each index, and then each step. Low implementation = 0 - 25%; partial = 25.1 - 74.9%; moderately high = 75 - 89.9%; high = >90%.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	The principal investigator then interviewed postpartum nurse managers by telephone over a 2-month period.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.19 D'Abbs et al., 2008

Publication	First author/Year/ Title	**(d'Abbs et al., 2008). Implementing a chronic disease strategy in two remote Indigenous Australian settings: a multi-method pilot evaluation**
	Peer reviewed	Yes
Context	**Country**	Australia
	Settings (urban/rural)	Primary health units at the regional health system level providing care to Indigenous communities
Background	**Aim/primary objective**	To conduct an evaluation of a framework developed for the North Queensland Indigenous Chronic Disease Strategy.
	Target population	Health centres serving Indigenous communities in North Queensland
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	The authors created their own framework to measure a process evaluation of implementation.
Methodological development steps		1. Framework is comprised of 2 components: Process evaluation and outcome evaluation (for the purposes of this review only process evaluation tools were used) 2. Process evaluation uses a System Assessment Tool (SAT) which was developed by adaptation of the Assessment of Chronic Illness Care Scale.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	7
	Nature e.g. exit interviews, checklist etc.	Semi-structured interviews with a knowledgably staff worker, as well as agency records and direct observation
	Structured/semi/open-ended questionnaire	Structured and unstructured

Contents	**Thematic areas and their operational definition given in the paper**	1) *Organisational influence* refers to the use of organisational influence to create a culture, and support organisational systems and mechanisms to promote safe, high quality care. Links within the community and with external (outside) services refers to the extent to which the primary health care service uses linkages with other community resources and service providers to promote health and enhance the effectiveness of the service. 2) *Self-Management Support* refers to health service systems that support development of awareness and skills for clients and families to take a major role in maintaining their health, managing health problems, and promoting a safe and healthy environment. 3) *Decision Support* refers to systems to ensure easy accessibility of evidence based information to inform decisions related to individual clinical care, and the promotion and protection of the health of the service population. 4) *Delivery System Design* refers to the extent to which the design of service infrastructure, work flow, staffing and other service delivery systems maximises the potential effectiveness of the health service. 5) *Information systems* refers to the quality and use of systems to generate and share information about the health service client population and individual clients, and the service processes and outcomes to inform decisions about effective service planning and operation. 6) *Integration* refers to how well the above components of the service work in a mutually supportive way that maximises the potential effectiveness of the health service.
Scaling	**(e.g. binary, likert scale etc)**	Scale 0-11
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	For the SAT: For each question, respondents choose one of three ratings based on a 0-11 subscale: 0–2 (little or no support); 3–5 (basic or intermediate support); 6–8 (advanced support); and 9–11 (fully developed support).
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	In two communities, a Diabetes Chart Audit and a Preventive Services Chart Audit was undertaken over 3 days in November to December 2005 by a registered nurse. The SAT was completed by a researcher acting as an external facilitator in about three hours in each community. Indicator assessment was completed by Queensland Health.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	Maternal Health and Child Health Systems Assessment Tool. Version 2.1 (Menzies School of Health Research, 2007).

2.20 Rosecrans et al., 2008.

Publication	First author/Year/ Title	**(Rosecrans et al., 2008). Process evaluation of a multi-institutional community-based program for diabetes prevention among First Nations**
	Peer reviewed	Yes
Context	**Country**	Canada
	Settings (urban/rural)	First Nations rural communities
Background	**Aim/primary objective**	To evaluate the implementation of the "teaching to prevent diabetes" community-based programme
	Target population	Community based programmes involved in diabetes programmes among Native North Americans
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	Developing a process evaluation plan for assessing health promotion program implementation: a how-to guide (Saunders et al., 2005); Process Evaluation for Public Health Interventions and Research (Linnan L and A., 2002).
Methodological development steps		1. Evaluation instruments developed based on previous work and lessons learned; 2. Standards for intervention delivery set for each component; 3. Process indicators set by agreement by the authors (e.g. fidelity % of minimum foods stocked/dose received/% of family pack cards completed and returned)
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	Various number of items depending on a component (e.g. school-based: 5 sub-items; stores: 5 sub-items; health and social services: 7 sub-items).
	Nature e.g. exit interviews, checklist etc.	Surveys, logs, interviews, checklists, semi-structures interviews
	Structured/semi/open-ended questionnaire	Structured
Contents	**Thematic areas and their operational definition given in the paper**	Five constructs: 1) 'fidelity'- quality of programme delivery and extent to which it is delivered as planned; 2) 'dose delivered' - number of units delivered by interventionists; 3) 'dose received'- extent to which target audience actively engages in and receives intervention activities; 4) 'reach' - amount of target audience that participates in the

		intervention; 5) 'context' - larger socio-political and environmental factors that may influence the intervention
Scaling	**(e.g. binary, likert scale etc)**	Scale (%)
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Implementation was rated 0-49% as "low", 50-74% as "moderate" and 75-100% as "high"
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	All forms were regularly sent by fax or hand-delivered to field supervisor. Process evaluators employed by the programme; research staff and investigators collected the data
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.21 Wilson et al., 2010

Publication	First author/Year/ Title	**(Wilson et al., 2010). Do intervention fidelity and dose influence outcomes? Results from the Move to Improve worksite physical activity program**
	Peer reviewed	Yes
Context	**Country**	USA/Canada
	Settings (urban/rural)	Worksites of a large retail organisation implementing the Move to Improve programme
Background	**Aim/primary objective**	To evaluate the implementation of a worksite physical activity program
	Target population	Office workers
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Research
Conceptual basis	**Name & reference of the theoretical framework**	Program Implementation in Preventive Trials (Durlak, 1998a).
Methodological development steps		3 questionnaires developed to assess implementation of active ingredients specific to 'site co-ordinators', 'team captains', and 'employees'. Questions asked about intervention participation, barriers, support and implementation
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	questionnaire for site co-ordinator = 17; questionnaire for team captain = 19; questionnaire for employee = 12
	Nature e.g. exit interviews, checklist etc.	Interviews, questionnaires by telephone
	Structured/semi/open-ended questionnaire	Structured
Contents	**Thematic areas and their operational definition given in the paper**	1)Intervention fidelity: degree to which the protocol was implemented as planned; 2)dose delivered: amount of the intervention delivered; 3)dose received: amount of the intervention received by the participants
Scaling	**(e.g. binary, likert scale etc)**	5 point scale (questionnaires for site coordinator/team captain) and yes/no questions

		(employee)
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Four rankings were averaged to derive an overall ranking for each site. Sites were further split at the median into two categories: high and low implementation.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Questionnaires were administered at the end of the intervention; site coordinators and team captains were interviewed at the work sites; employees completed the questionnaire via phone.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	The level of implementation was associated with the degree of change in physical activity: Mean 5 5.4 versus 2.2; Chi-square = 4.9, df = 1.
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.22 Edward et al., 2011.

Publication	First author/Year/ Title	**(Edward et al., 2011). Configuring Balanced Scorecards for Measuring Health System Performance: Evidence from 5 Years' Evaluation in Afghanistan.**
	Peer reviewed	Yes
Context	**Country**	Afghanistan
	Settings (urban/rural)	Primary health care services/systems
Background	**Aim/primary objective**	To examine trends in health system performance indicators and to test a balanced scorecard (BSC) as a performance management tool
	Target population	Health facilities (e.g. hospitals, health centres, mobile clinics) in 28 provinces of Afghanistan
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	The balanced scorecard: translating strategy into action. (Kaplan and Norton, 1992).
Methodological development steps		1. A BSC was developed based on standards instituted in the Basic Package of Health Services. 2. Traditional BSC quadrants were modified to include 6 domains with 29 core performance indicators, designed by a multidisciplinary team of government donor, and NGO stakeholders.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	29 Some items are tools (e.g. outpatient visit concentration index) which are aggregated to individual indicators
	Nature e.g. exit interviews, checklist etc.	Patient observations and exit interviews with patients and providers
	Structured/semi/open-ended questionnaire	Structured
Contents	**Thematic areas and their operational definition given in the paper**	1)Patient and community perspectives: patient satisfaction and engagement of community councils; 2)staff: workforce capacity, management and satisfaction; 3)capacity for service provision: system preparedness based on BPHS standards for staffing, equipment, essential commodities, and infrastructure; 4)financial systems: user fees; 5) overall vision: equity factors.
Scaling	**(e.g. binary, likert scale etc)**	Scale %

Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Each indicator was converted to a % score (0 to 100). Provincial performance was applied to set the benchmarks and provinces were categorised into quintile groups on the basis of performance with the top and bottom quintiles illustrating upper and lower benchmarks. Fourteen of the 29 indicators were indices, created from an aggregate set of performance indicators.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Patient observations by a five-member survey team were conducted on five children under 5 y, and five patients above 5 y, selected by systematic random sampling using a sampling interval on the basis of utilization rates between in 2004-2008
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	The developed scorecard has been adopted by the Ministry of Public Health. Similar tool was used to measure health system performances both in low and high income countries (Khan et al., 2012, Lupi et al., 2011, ten Asbroek et al., 2004).

2.23 Perez et al., 2011

Publication	First author/Year/ Title	**(Perez et al., 2011). Process-oriented fidelity research assists in evaluation, adjustment and scaling-up of community-based interventions**
	Peer reviewed	Yes
Context	**Country**	Cuba
	Settings (urban/rural)	Programmes to assess the effectiveness of community participation in the control of dengue fever
Background	**Aim/primary objective**	To assess fidelity and reinvention in the implementation of a community-based control strategy for *Aedes aegypti* (dengue fever); to understand difficulties and barriers faced during implementation; to reveal information on the feasibility of implementing the strategy on a larger scale
	Target population	Ministry of Health, Cuba and community-based programmes involved in dengue fever control
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	Translating research into practice: the dissemination and initial implementation of an evidence-based HIV prevention program (Rebchook et al., 2006); Diffusion of innovations (Rogers, 2003).
Methodological development steps		1. Four major components identified for the evaluation. 2. Three-step assessment developed: individual evaluation by professionals; discussion of these assessments by a broader group of experts involved in different stages of the strategy's implementation; refining the assessment within a participatory process evaluation workshop
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	Correlation among the components was >0.70 on Spearman Rank coefficient
Structure	**No. of items**	23
	Nature e.g. exit interviews, checklist etc.	Analysis of documentations, individual assessments with management, semi-structured interviews
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1)Organisation and management; 2)capacity building; 3)community work; 4)surveillance. The detailed definitions and descriptions of the thematic areas are provided in Appendix

		3.17, Table 26.
Scaling	**(e.g. binary, likert scale etc)**	Scale of four: Implemented, Modified, Not Implemented, Added
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Major components scored 0 if not implemented, 1 if modified and 2 if fully implemented
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Evaluations were carried out by analysis of proceedings, minutes and process documentation forms; six professionals further assessed if the strategy was implemented as specified; trained researchers conducted qualitative interviews
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	The study measure immunisation coverage as their outcome and have concluded that the percentage of districts with the coverage ≥80% increased as the overall RED implementation score increased.
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.24 Ryman et al., 2011

Publication	First author/Year/ Title	**(Ryman et al., 2011). Implementation of the reaching every district (RED) approach: experience from North Sudan**
	Peer reviewed	Yes
Context	**Country**	North Sudan
	Settings (urban/rural)	Immunisation services utilising the Reaching Every District (RED) Approach
Background	**Aim/primary objective**	To assess the extent and quality of implementing RED in North Sudan and it's impact on immunisation coverage
	Target population	70 districts of North Sudan providing immunisation to every child
	Primary discipline (health, education, etc)	Health services
	Research or service delivery programme	Service delivery programme
Conceptual basis	**Name & reference of the theoretical framework**	The Reaching Every District approach (World Health Organization, 2008).
Methodological development steps		1.Two types of data collection worksheets were developed: 1)for self-completion by the Locality Officers (district routine immunisation managers); 2)to guide a peer in reviewing district level immunisation documents. 2.Indicators for each thematical areas were weighted by an expert 3.A rating system from general assessment and peer-reviewed sheets was developed.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	No psychometric characteristics but validation of data collected from district staff through worksheets found a high level of agreement.
Structure	**No. of items**	29
	Nature e.g. exit interviews, checklist etc.	Worksheets for district-level staff in all study districts, semi-structured interviews in four selected districts, district-level documents and data review (e.g. monitoring charts, micro-plans)
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1)Re-establishing outreach : percentage of outreach, mobile and fixed sessions held; supportive supervision: number of supervisory visits and reports, quality of supervision records; 3)Monitoring and use of data for action: monitoring charts available and their quality, percentage of sites sending reports, number of review meetings; 4)strengthening planning and management of resources through micro-plans: vaccine, syringe, vaccine card, telly sheet stock-out, staff training, equipment functioning; 5)increasing community links: social mobilisation activities, community groups involved

Scaling	**(e.g. binary, likert scale etc)**	10 point scale
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	10=fully implemented; Implementation score was calculated by summing five component scores. "implementation groups" below 5.28 score were defined as low, 5.28/6.31 - medium, more than 6.31 - high.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Locality Officers in each district completed the worksheets between February and March 2007. A peer locality officer led a review and rating for the supplemental documents.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.25 Sogarwal., 2011

Publication	First author/Year/ Title	**(Sogarwal and Bachani, 2011). Process Evaluation of Community Care Centers Providing Care, Support and Treatment to People living with Human Immunodeficiency Virus in India**
	Peer reviewed	Yes
Context	**Country**	India
	Settings (urban/rural)	Community Care Centres (CCC) for people living with HIV
Background	**Aim/primary objective**	To present the process evaluation of CCC and identify gaps/constraints to implementation of the scheme.
	Target population	Inpatient/outpatient beneficiaries, caregivers and service providers of CCCs for HIV/AIDS that have been running for at least 1 year
	Primary discipline (health, education, etc)	Community health (HIV/AIDS)
	Research or service delivery programme	Service Delivery Programme
Conceptual basis	**Name & reference of the theoretical framework**	National Operational Guidelines for CCCs under NACO(National AIDS Control Organization et al., 2007).
Methodological development steps		One quantitative and two semi-structured tools were designed. For the quantitative tool: 1) the data collection team recorded the information/data through direct observation. 2) For each attribute the team recorded the most appropriate code on the basis of predetermined options. 3). The final score for each CCC was computed.
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	18
	Nature e.g. exit interviews, checklist etc.	Direct observations, questionnaires and individual interviews, checklists and semi-structured tools.
	Structured/semi/open-ended questionnaire	Structured and unstructured
Contents	**Thematic areas and their operational definition given in the paper**	1)Physical infrastructure: accessibility, physical ambience and hygiene, adequacy; Systems: adequate and functional equipment, adequacy of supplies and waste management infection control measures being followed, linkages, referrals, and coordination; 2) human resources: adequacy and training of staff; 3) services: services delivery, support services, level of patient satisfaction, bed utilization; 4) financial and inventory management: maintenance of accounts , procurement and supplies, expenditure and audit; 5)monitoring

		and evaluation: regularity in updates, record keeping
Scaling	**(e.g. binary, likert scale etc)**	% scores and likert scales (very satisfied to very dissatisfied).
Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	Tool 1 scores were added together, and information from tools 2 and 3 were used to produce a final score. This determined the future of the programme: a Grade A (≥80%) = Contract continued; Grade B (70-79) = Contract continued with recommendations; Grade C (60-69) = Temporary extension with major recommendations; Grade D (≤60%) - Contract cancelled.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	Data were collected between January and March 2010. Each process evaluation data collection team visited the centre and observed the physical infrastructure.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

2.26 Yumo et al., 2011

Publication	First author/Year/ Title	**(Yumo et al., 2011). WHO recommended collaborative TB/HIV activities: evaluation of implementation and performance in a rural district hospital in Cameroon**
	Peer reviewed	Yes
Context	**Country**	Cameroon
	Settings (urban/rural)	Rural district hospital initiating TB/HIV activities
Background	**Aim/primary objective**	To assess implementation and performance of recommended collaborative TB/HIV activities
	Target population	Medical and para-medical staff involved in TB and/or HIV programmes
	Primary discipline (health, education, etc)	Health
	Research or service delivery programme	Service Delivery Programme
Conceptual basis	**Name & reference of the theoretical framework**	A guide to monitoring and evaluation for collaborative TB-HIV activities (World Health Organization, 2009).
Methodological development steps		1. Core indicators were identified using the WHO recommended TB/HIV activities. 2. The performance of implemented activities were calculated
Psychometric characteristics	**(Tool development methodologies sometimes include information on reliability, validity etc)**	None
Structure	**No. of items**	8
	Nature e.g. exit interviews, checklist etc.	Interviews, a review of registers and drug-stock cards, a questionnaire
	Structured/semi/open-ended questionnaire	Structured
Contents	**Thematic areas and their operational definition given in the paper**	Areas to reduce burden of HIV/TB: HIV counselling and testing for TB patients; HIV prevention among TB patients; preventive therapy to TB/HIV co-infected patients; HIV care and support to TB/HIV patients; Antiretroviral therapy to TB/HIV co-infected patients; TB screening and diagnosis in HIV positive patients; preventive therapy for HIV patient with latent TB; TB prevention control in the hospital.
Scaling	**(e.g. binary, likert scale etc)**	Binary (yes/no)

Scoring	**(e.g. categorisation of total scores of a given construct or overall tool)**	The score = a rate calculated by dividing the total number of recommended activities effectively implemented in the hospital by the total number of recommended activities for the operational level.
Data source	**Primary**	Primary
	Secondary	
Conduct	**Brief description of data collection process (by whom, timing)**	A cross-sectional study was conducted in July 2009, records of all patients treated in the hospital for TB from 2003-2008 were reviewed.
Associational information	**Tool scores association with the study outcome (measure and strength of association)**	None
Evidence of further use of the tool	**Reference if other studies have used same or adopted version of the given tool**	None

Table 3. Glossary of implementation strength.

Term	Definition/description	Source
Implementation strength	Implementation strength refers to the *quantity* of a program strategy that is carried out at the field/population level and incorporates some elements commonly considered as part of the *quality* of service delivery as well…The measurement of implementation is intended to reflect the amount of the program that is *delivered*.	(Bryce et al., 2011)
	The aggressiveness of the States efforts to deliver the programme Wide geographical spread of company marketing activities, targeting all market segments would indicate greater implementation strength	(Gold et al., 1993) (Eikeland, 2007)
	"Programme strength" - level of programme effort to improve maternal and newborn health. "Program strength" ... refers to the dose, duration, specificity, and intensity of intervention...[and answers a question] "How many hours of service are needed to develop a clinically meaningful level of change in clients?"	(Ross and Begala, 2005) (Summerfelt, 2003)
Implementation intensity	Quantity and depth of implementation activities "Programme intensity" = 'dose' - the intensity (frequency and duration of interventions) and complexity (number of interventions) of the program. "District-level programme intensity" –the annual number of people newly starting ART per capita in the district, and the annual number of people starting ART per PEPFAR-supported clinic in the district.	(Pearson et al., 2005) (Riegel et al., 2010) (Bendavid et al., 2012)
	How well a proposed program of intervention is put into practice.	(Durlak, 1998b)
Implementation rate	The number of weeks in which a school successfully completed the 12 implementation tasks required to reach full implementation.	(Buzhardt et al., 2006)
Implementation quantity	How much of the intervention was delivered; the "dosage of the intervention".	(Hebert and Veil, 2004, McGraw et al., 2000)
	Dose delivered (the amount of the intervention delivered) and dose received (the amount of intervention received by the participants).	(Wilson et al., 2010)
	The extent and rate of clinics' adoption of key program components, the amount and nature of assistance received from staff, the volume and type of cholesterol management services provided to patients.	(Glanz et al., 1992)

	Length of time in treatment, number of sessions attended, number of people completed the treatment programme... the number of services received multiplied by the duration of the service.	(Orwin, 2000)
	'Dose delivered'- the number of units delivered by interventionists; 'dose received' - the extent to which the target audience actively engages in and receives intervention activities; 'reach', the amount of target audience that participates in the intervention.	(Rosecrans et al., 2008)
Implementation degree/ level/extent	The degree to which the intervention can be adapted to fit the local context, the strength and quality of the evidence supporting the intervention, quality of design and packaging and cost. The extent to which the clientele using the services corresponds to the clientele initially targeted; if the services delivered correspond to those planned; if resources planned were effectively made available; and if the delivery procedure corresponds to the one initially defined.	(Damschroder and Hagedorn, 2011) (Hebert and Veil, 2004)
	The extent to which the institution employs the Continius Quality Improvement concept and tools in its management protocol	(Lee et al., 2002)
	The extent to which a program offers services meeting the requirements of selected key elements (e.g. number of key elements implemented; programs with higher intensity had increased frequency of completed home visits, the hours they provided care in their centers, the frequency with which they offered group socialisation activities)	(Paulsell et al., 2002)
	The degree of a programme implemented as planed	(Orwin, 2000)
	The extent of implementation: the proportion of the target population for a required element who are actually affected by that element's implementation, or the frequency with which a required event or activity occurs.	(Smith et al., 1977)
	The implementation level was determined as the implementation score of the recommended activities. The score = a rate calculated using as numerator the total number of recommended activities effectively implemented and as denominator the total number of recommended activities for the operational level. The extent to which structural and interactional critical components are present or not	(Yumo et al., 2011) (Cassata-Widera et al., 2011)
Implementation quality/ fidelity components	Fidelity - the extent to which delivery of an intervention adheres to the protocol or program model originally developed. The components of Implementation fidelity: adherence, dosage, quality of delivery, participant responsiveness and	(Dusenbury et al., 2003, Mowbray et al., 2003, Carroll et al.,

	programme differentiation	2007)
	Fidelity – the degree to which a programme's design was followed.	(Vinson et al., 2001)
	Dosage—providing sufficient exposure to the program, (2) adherence—following program methods and completing its delivery as outlined in a manual or curriculum guide, (3) quality of process—engaging students through their active participation, and (4) adaptation— modifying the program to meet developmental and cultural needs.	(Dusenbury et al., 2003)
	The degree to which group leaders deliver the intervention competently and according to protocol. The determination of how well an intervention is implemented in comparison with the original program design during efficacy and/or effectiveness study	(Breitenstein et al., 2010) (O'Donnell, 2008)
Questions to assess implementation at programme level	"(a) were facilities acquired as intended? (b) were target sample sizes achieved? (c) were specific subgroups recruited as intended (for example, women)? (d) were planned service components successfully implemented? (e) was the intensity or "dose" of the intervention consistent with the program model? (f) was the intervention stable once implemented? and (g) were systems-level objectives met?	(Orwin, 2000)
Implementation success	The presence of the innovation as intended in the delivery context	(Smith, 2010)
	Implementation success is determined by a large-scale implementation effort in six areas 1. Well-defined key result areas and goals; 2. Well-defined objectives; 3. Well-defined measurement process; 4. Well-established reviews; 5. Responsibilities defined and visible; 6. Continuous improvement processes exist Two factors determine a success of reengineering projects: breadth and depth. 1) The process should be broadly redesigned 2) redesign should penetrate six depth levers: roles and responsibilities; measurements and incentives; organizational structure; information technology; shared values; and skills.	(Hacker and Washington, 2004) (Hall et al., 1994)

Appendix 2. Review tools

Table 4. Search strategy

Database	Set	Searches
MEDLINE	1	Measure* implementation
	2	Limit 1 to English
	3	Limit 2 to Humans
	4	Evaluate* implementation
	5	Assess*implementation
	6	Examine* implementation
	7	Implementation intensity
	8	Implementation strength
	9	Implementation extent
	10	Implementation degree
	11	Implementation level
	12	Implementation rate
	13	Process evaluation AND implementation
		Strategy: 2 and 3 and (1 or 4 or 5 or 6 or 7 or 9 or 10 or 11 or 12 or 13)
Embase	1	"Measuring implementation" or "measure implementation" or "implementation measurement"
	2	Limit 1 to English
	3	Limit 2 to Humans
	4	"Evaluating implementation" or "implementation evaluation" or "evaluate implementation"
	5	"Assessing implementation" or "implementation assessment" or "assessing implementation"
	6	"Examine implementation" or "examining implementation"
	7	"Implementation intensity"
	8	"Implementation strength"
	9	"Implementation extent"
	10	"Implementation degree"
	11	"Implementation level"
	12	"implementation rate"
	13	"Process evaluation implementation"
		Strategy: 2 and 3 and (1 or 4 or 5 or 6 or 7 or 9 or 10 or 11 or 12 or 13)
ERIC	1	"Measuring implementation" or "measure implementation" or "implementation measurement"
	2	"Evaluating implementation" or "implementation evaluation" or "evaluate implementation"
	3	"Assessing implementation" or "implementation assessment" or "assessing implementation"
	4	"Examine implementation" or "examining implementation"
	5	"Implementation intensity"
	6	"Implementation strength"
	7	"Implementation extent"
	8	"Implementation degree"
	9	"Implementation level"
	10	"implementation rate"
	11	"Process evaluation implementation"
		Strategy: 1 or 2 or 3 or 4 or 5 or 6 or 7 or 9 or 10 or 11

Table 5. Main websites used in grey literature search.

Agency	Website
The Bill & Melinda Gates Foundation	http://www.gatesfoundation.org/Pages/home.aspx
Doctors without Borders (Medecins sans Frontiers)	http://www.doctorswithoutborders.org/
The Kaiser Family Foundation	http://www.kff.org/
Open Society Foundations	http://www.soros.org/
Partners in Health	http://www.pih.org/publications/
Popline	http://www.popline.org/
Public Health Institute	http://www.phi.org/resource_library/index.html
Google Scholar	http://scholar.google.co.uk/
Social Policy and Practice	http://ovidsp.tx.ovid.com/sp-3.5.1a/ovidweb.cg i
United Nations	http://www.un.org/en/ http://www.unfpa.org/public/ http://www.unaids.org/en/ http://www.undp.org/content/undp/en/home.html, http://www.unodc.org/
World Health Organization	http://www.who.int/en/

Table 6. List of experts contacted.

Murdo Bijl, Director, Health Connections International, Amsterdam, Netherland

Dr. Jennifer Bryce, John Hopkins Bloomberg School of Public Health, Baltimore, USA

Dave Burrows, Director, AIDS Projects Management Group, Australia

Sapna Desai, Director, Women's Health and Health Insurance Study, Self Employed Women's Association, Family Health International, UNFPA, New Delhi Area, India

Professor Andy Haynes, London School of Hygiene and Tropical Medicine, London, UK

Dr. Gregory B. Teague, University of South Florida, USA

Professor Cesar G. Victora, John Hopkins Bloomberg School of Public Health, Baltimore, USA

John Walker, Head of Group Commercial Development at Home Retail Group PLC, London, UK

Table 7. Data extraction form for the tools measuring implementation strength.

Publication	a) Author/s & Year/ title b) Peer reviewed
Context	a) Country. b) Settings
Background	a) Aim / primary objective b) Target population c) Primary discipline (health, education etc) d) Research or service delivery programme
Conceptual basis	Name & reference of the theoretical framework
Methodological development steps	Enlist steps of told development or tool development
Psychometric characteristics	(Tool development methodologies sometime include information on reliability, validity etc)
Structure	c) Structured/ semi/open ended questionnaire etc a) No of items b) Nature e.g. exit interviews, checklist etc
Contents	Enlist thematic areas and their operational definition given in the paper
Scaling	(e.g., binary, likert scale etc)
Scoring	(e.g., categorization of total scores of a given construct or overall tool
Data source	a) Primary b) Secondary (source e.g. existing record, MIS)
Conduct	a) Brief description process data collection process b) by whom c) Timing
Associational information	Tool scores association with the study outcome (measure and strength of association)

Appendix 3. Studies' descriptions

In this section information for each study on scaling and scoring is provided, and the content of the tools is described. Studies are presented in chronological order.

(Smith et al., 1977) developed an implementation rating instrument to measure implementation of the Project Developmental Continuity. The instrument consisted of two types of scales: objective scale and judgemental scale (in total, 350 scales). An example of judgemental scale is provided in Appendix 3.1, **Error! Reference source not found.**. These 350 scales were grouped into ubcomponent clusters. To produce a single subcomponent score, scores on items within the subcomponents were averaged. The overall score was an average of subcomponent scores.

The content of the instruments covers seven areas and is based on the Guidelines and the Project Developmental Continuity programme requirements: 1) administration (5 subcomponents); 2) education (4 subcomponents); 3) bilingual/bicultural and/or multicultural (4 subcomponents); 4) handicapped (4 subcomponents); 5) parent involvement (2 subcomponents); 6) developmental support services (4 subcomponents); and 7) pre-service and in-service training (7 subcomponents). The whole list of subcomponents and their descriptions is provided within the Smith et al., report as a separate file (pp.239-253).

(Rubin et al., 1982) introduced the Field-based Implementation Rating Scale to measure implementation of field-based educational programmes. The scale includes specifications for each criterion or programme attribute which range "from gross indicators of incidence of activities to systematic observation requiring subjective judgement… the elements of the rating scale include columns containing descriptions of the activities and attributes associated with the educational program, criterion levels for each attribute, data sources and monitoring documenting activities for measurement of each attribute, reports of the level of implementation, and an indication of whether the criterion level has been met (Appendix 3.2). The level of implementation is shown in percentages. No composite scores are built.

The whole tool is not provided, so it was not clear whether the items were categorized in any way and what type of items were included. However, the authors mentioned that the specifications for the attributes were identified by programme staff, community stakeholders, and through modifications and verifications of the initial programme model.

(Glanz et al., 1992) evaluated the implementation of a cholesterol management programme in family physician offices. Programme staff rated the implementation of each essential component (see below) on a scale 1-5 where 1='not at all in place' or not implemented, 5='completely in place' or fully implemented. The quantity of implementation was also assessed by counting the number of patients screened and given counselling if they had elevated cholesterol levels.

The key components of the tool were identified using the National Cholesterol Education Program guidelines for evaluation and treatment of elevated cholesterol in adults. The 14 components were further categorized based on the related literature in four categories: screening, physician message, counselling, and tracking and monitoring (Glanz, 1985, Snetselaar, 1989). The whole list of essential components is provided in Appendix 3.3.

(Gold et al., 1993) evaluated strength of state implementation efforts in expanding Medicaid (a large federal-level programme) eligibility criteria for pregnant women. The instrument comprised 48 items and measures in four domains. Each domain and items within domains are weighted on 0-100 percentage scale. The overall Index of Strength of Implementation is a sum of scores of the ranked items with '0' indicating the least overall effort and '100' – the most possible effort. Each state is then further ranked with the 10 top-ranking states receiving rank 1, and the lowest ranking 10 receiving a rank 5.

Four major domains were identified for the instrument: Eligibility and Enrolment (has three components and 16 items), Enhanced services (two components and 8 items), Outreach Activities (two components and 22 items), Reimbursement Generosity (two components). These domains reflected the four main aspects of the Medicaid expansions. The items included: availability and types of services, costs, geographic coverage, target population, level of reimbursement, etc. The full list of items with assigned weights is provided in Appendix 3.4. The instrument was developed using different sources (surveys, published literature, records) by the Alan Guttmacher Institute.

(Shortell et al., 1995) assessed the degree of quality improvement implementation in 61 U.S. hospitals. The implementation was measured by six scales (one scale for each of the six domains). It is not clear, however, how scales were built and how the degree of implementation was calculated.

The six domains of the tool to measure implementation degree were based on the Malcolm Baldrige National Quality Award criteria and included: leadership, information and analysis, human resources management, quality management, and strategic quality planning. The whole content of the tool (e.g., items in each domain) are not provided.

(Bond et al., 1997) developed a brief 15 items, 5 point scale with Staffing (3 items), Organization (9 items) and Services (3 items) subscales to measure the extent to which vocational programmes for people with severe mental illness followed the Individual Placement and Support (IPS) model of Supported Employment (SE). The structure of the scale was based on two prior implementation/ fidelity scales in the mental health field.

The content of the items was taken from the IPS programme manual, brainstorming sessions among experienced implementers of this model, and SE literature. The whole scale structure and content is available within the article.

(Teague et al., 1998) built their fidelity implementation scale to assess a complex community-based service, Assertive Community Treatment (ACT). The structure of the scale was based on previous research in the same area (McGrew et al., 1994, Teague et al., 1995). The 5-point scale contains programme items and anchors for each point on a rating scale (Appendix 3.5, Table 13). The anchors were defined through published norms of the programme, published literature, experts' surveys, and authors' consensus. No composite score was calculated but means of items and overall scores were provided.

The items were comprised of elements of programme structure and processes and categorized in three domains: human resources (composition and structure of staffing), organizational boundaries (programme responsibility and relationships with other components) and nature of services (range, nature of services, overall treatment approach). The detailed content under each domain is provided

in Appendix 3.5, **Error! Reference source not found.**. To identify model components, literature escribing the model was searched, previous work in the same field was analysed and experts' opinion was used.

(Orwin, 2000) assessed treatment fidelity and programme implementation in multisite substance abuse services by employing various qualitative and quantitative tools: implementation histories, implementation scales, mapping and logic models. For example, for overall intervention strength, the formula was created which represented "the number of services received multiplied by the duration of the service". The formula for overall intervention strength calculation is provided as a separate file within the article (p.S327).

Although the main instrument for the data was the Quarterly Report Form, it was not clear how it was constructed and what exact content it had.

(Vinson et al., 2001) evaluated implementation fidelity of the system-of-Care model in 27 communities in the US and Hawaii. The implementation index was developed and contained 100 questions in two domains and 16 attributes. Each question was measured on a 4-point scale: 0=absence or minimal presence, 3=mature and systematic presence. Each attribute was provided a score after the scales were averaged. A total score was calculated by combining the scores of each attribute.

Two major domains were evaluated: infrastructure (10 attributes and 50 items) and service delivery (6 attributes and 50 items). Infrastructure includes organizational arrangements and processes, and service delivery "is application of system-of-care principles to the direct provision of care to children and families." The detailed description of each attribute is provided in Appendix 3.6. The content was developed and refined after a literature search, experts (researchers, practitioners, family members) feedback, and pilot-testing.

(Bonomi et al., 2002) tested the Assessment of Chronic Illness Care (ACIC) tool which evaluated strengths of delivery of care performance according to the Chronic Care Model. The tool is a 0-11 point scale, where 0-2 is categorized as little or no support to the model, 3-5 – basic or intermediate support, 6-8 advanced support, and 9-11 optimal support. A description is provided for each support level for each of the 28 components covering the six areas of the Chronic Care Model. For example, the item 'Practice Team Functioning' is considered to be Limited if it "*is not addressed*" and Full if it *"it is assured by teams who meet regularly and have clearly defined roles including patient self-management education, proactive follow-up, and resource coordination and other skills in chronic illness care"*. Scores for each subscale are summarized and averages presented.

The tool content and items categorization were based on specific interventions and concepts of the Chronic Care Model, literature search, expert panels and the adapted standard for organizational excellence, the Malcolm Baldrige National Quality Award Criteria(Shortell et al., 1995, U.S. Chamber of Commerce, 1993). 28 items were grouped in six domains which reflected six areas of the Chronic Care Model: organization of the health care delivery system (leadership, goals, strategies, incentives/regulations and benefits), community linkages (links to outside resources, community organizations, and regional health plans), self-management support (assessment of needs and activities, concerns of patients and families, behaviour change interventions and peer support, self-management support), decision support (evidence-based guidelines, specialist involvement, provider

education, informing patients about guidelines), delivery system design (functioning, leadership, appointment systems, follow-up planned visits, continuity), and clinical information systems (registry, reminders, feedback, treatment plans, information).

The whole tool with definitions of each item according to four levels is provided as a separate file Appendix within the article (pages 811-820).

(Lee et al., 2002) assessed the extent of continuous quality improvement (CQI) implementation in Korean hospitals. The degree of CQI implementation was measured by a questionnaire comprised of seven dimensions and 43 items using 5-point scale (1=strongly disagree/5=strongly agree). The mean score across seven dimensions were calculated at 67 hospitals.

The dimensions/domains of the instrument were based on the Malcolm Baldrige National Quality Award Criteria (US, Chamber of Commerce, 1993) and included: leadership, strategic planning, customer satisfaction, information and analysis, human resources management, quality management, and organizational performance results. The article does not specify items included under each domain. (Although not specific to this study, this tool can be viewed at http://www.nist.gov/baldrige/publications/upload/2011_2012_Business_Nonprofit_Criteria.pdf)

(Paulsell et al., 2002) assessed the extent of programme implementation and the quality of key child development services in 17 research programmes. The authors have developed scales for each program element to measure the degree of implementation which consisted of five levels of implementation, from minimal implementation (Level 1) to enhanced implementation (Level 5).

The tool content was based on the programme requirements, Head Start Program Performance Standards, and the Early Head Start grant announcement, and included three major program areas: (a) early childhood development and health services, (b) family and community partnerships, and (c) program design and management. The whole tool contents is available in Appendix 3.8.

(Cooley et al., 2003) designed the Medical Home Index to measure implementation of the Medical Home model. The scale includes four levels of achievement of the particular item with each level scored as 'partial' or 'complete' representing 'some activity within level' or 'all activity within level' respectively. The 4 levels represent: basic paediatric care (level 1) through responsive care (level 2), proactive care (level 3), and comprehensive care (level 4). The structure of the scale was adapted from the ACIC tool with the addition of Partial/Complete scoring for each of four scale levels (Appendix 3.9).

The raw summary scores were transformed to a 0-100 scale, and mean summary scores were calculated.

The instrument' domains reflected practice activities identified as critical to the quality of care of a Medical Home Model. They included organizational capacity, chronic condition management, care coordination, community outreach, data management and quality improvement. The domains encompassed characteristics from the Medical Home Model definition and were refined after discussions with federal and national leaders, Medical Home experts and representatives of NGOs. Items under domains are provided in Appendix 3.9., **Error! Reference source not found.**. The

uthors, however, do not specify how themes under the domains were identified, and how level definitions were constructed.

(Hacker & Washington, 2004) evaluated large-scale organizational change (how much the organization implemented the change) in 16 ministries in public civil service in Botswana using a seven point scale. Each point on the scale provided a short description: 1-implementaiton not started in the organization, 2-at least one department partially implemented in the organization, 3-some departments partially implemented in the organization, 4- partial implementation across all departments in the organization, 5- at least one department fully implemented in the organization, 6- some departments fully implemented in the ministry, 7- implementation complete across all ministry. Average scores for each of six domains (see below) are provided.

The authors identified six areas (which included 42 items) to be measured based on literature searches and their experience: 1. well defined key result areas and goals; 2. well defined objectives; 3. well defined measurement process; 4. well established reviews; 5. and 6. Continuous improvement processes exist. The whole tool content is available in the Appendix 3.10.

(Hebert and Veil, 2004), evaluated processes of implementation of Integrated Service Delivery systems. A percentage scale was used to measure implementation of each component. An implementation degree score was provided for each indicator (in percentage) as well as an average, combined score for each domain.

The measurable indicators of implementation for each domain and their objectives were set by two committees consisting of meeting researchers, policy-makers, managers and clinicians involved in the project. The committees provided percentage weights for each indicator and the domain based on their relative importance. The domains were based on a PRISMA model which included: 1) coordination between decision-makers and managers at the regional and local level, (2) a single entry point, (3) a case management process, (4) an individualized service plan, (5) a single assessment instrument coupled with a management system based, and (6) a computerized clinical chart. The items of the main six components and their weight are listed in Appendix 3.11.

(Kansanaho et al., 2005) assessed an implementation of a national project to promote professional strategy in Finnish community pharmacies. The implementation rate was measured by a questionnaire which consisted of 16 action statements, which were answered using a 5-point Likert scale (strongly agree/strongly disagree). The items were then recalculated (reverse coded): ''well implemented'' (3), ''implemented quite well'' (2), ''implemented to some extent'' (1) ''I cannot say'' (0) and ''not implemented at all'' (0). The summary of the scales were calculated, and categorized in three groups: a low implementation rate between 0 and 15 points, moderate implementation rate between 16 and 31 and high implementation rate between 32 and 48 points. The results were provided in percentages.

(Pearson et al., 2005) measured 42 organizations' implementation intensity and depth of Chronic Care Model by counting total changes made by the organizations. The depth ratings were presented as percentages of the maximum rating possible in each category.

The content of the tool was based on the Chronic Care model and the ACIC tool and consisted of six domains and 23 components (see Appendix 3.13).

(Grizzard et al., 2006) assessed implementation of the Ten Steps to Successful Breastfeeding programme in hospitals in Massachusetts, USA. The instrument was a cross-sectional survey consisting of 55 yes/no and 1-5 points Likert scale format questions (Appendix 3.14.). The implementation was quantified using a two-stage algorithm which converted responses to 0-100% scale. The scores were further categorized into four descriptive groups: low implementation (0–25%), partial implementation (25.1–74.9%), moderately high implementation (75–89.9%), and high implementation (90% and above).

The content was built on the WHO/UNAIDS Ten Steps to Successful Breastfeeding model, and had ten domains. Each domain consisted of two or three indices (sets of questions), reflecting key aspects of each step. The indices and questions were developed after a literature review and 25-person committee approval (nurses, lactation consultants, physician, childbirth educators). The instrument was piloted, and further revised by the researchers.

(d'Abbs et al., 2008) provided a process evaluation of implementation of management of chronic diseases at the primary health centre level in indigenous communities. The process evaluation was measured by the System Assessment Tool, an adaptation of the ACIC tool described above(Bonomi et al., 2002). The tool is 0-11 scale with descriptive anchors provided for each indicator and each point on the scale. Mean scores are calculated for each of seven domains. The scale levels are grouped in the following way: 0–2 (little or no support); 3–5 (basic or intermediate support); 6–8 (advanced support); and 9–11 (fully developed support).

The content of the tool is organized around seven dimensions (domains) of health service systems (Appendix 3.15.). The authors do not provide full tool content. We provide one of the versions as a separate file found at:

http://www.health.gov.au/internet/h4l/publishing.nsf/Content/01613125541AB991CA2572F4001A61D4/$File/Maternal%20and%20Child%20Health%20Systems%20Assessment%20Tool.pdf

(Rosecrans et al., 2008) measure fidelity, reach and dose (dose delivered and dose received) of implementation of diabetes prevention intervention among First Nation populations. A simple percentage scale was used (e.g., % of family pack cards returned, % minimum foods stored, % placement of shelf labels)) measuring fidelity, dose and reach in schools, at the stores and health and social services. Data was collected using multiple methods such as interviews, logs, check-lists and recordings. Components were: fidelity, dose delivered, dose received, reach, and context. The indicators were further categorized into low (0–49%), moderate (50–74%) or high (75–100%) levels. The indicators for some components are provided within the article (pp 277-279).

(Wilson et al., 2010) measured the implementation of the worksite physical activity programme at 16 worksites in the US and Canada. A questionnaire with 5-point scales was used to assess implementation (the instrument is not available). Average scores of four themes were calculated. The 26 sites were then "split at the median" to produce two groups: high and low implementation.

The 'Active ingredients' of the tool reflected the goals of the programme and included five areas: personal goal setting, team competition, environmental supports, management support, and environmental prompts. Within each components fidelity, dose delivered (e.g., % of participants receiving a manual) and dose received (% of participants using the manual) were measured. These

themes were not used, however, in calculation of overall implementation scores (see above). Instead, the authors used responses to questions in four areas to rank the sites 1) implementation barriers 2) participation levels 3) support for intervention activities and 4) overall implementation. Because a questionnaire was not provided in the article, it was not clear whether above five areas is included in these four themes.

(Edward et al., 2011) implemented a balanced scorecard (BSC) to evaluate primary health care services performance. The structure and partially the content of the tool were adapted from a Kaplan and Norton instrument, which measured industry performance (Kaplan and Norton, 1996). Each indicator was scored on 0-100 percentage scale. An overall national score for each indicator was calculated as a median of weighted provincial mean scores.

The tool included 29 indicators, some of which are aggregates of individual indicators (e.g., patient perception of quality index). The indicators were categorized in six domains: ''Patient and Community Perspectives'' (patient satisfaction and community council's activities), ''Staff domain'' (workforce capacity, management, salary and satisfaction), "Capacity for Service Provision'' (staffing, equipment, essential commodities, and infrastructure), ''Service Provision'' (clinical quality), ''Financial System'' (user fees) ''Overall Vision'' (equity factors). The BSC was designed by the representatives from the government, donor, NGOs and other stakeholders. The whole tool is provided in the Appendix 3.16.

(Perez et al., 2011) assessed process-oriented fidelity of the implementation of a community-based control strategy for Aedes aegypti control in 16 areas in Havana. They have used a three-point scale which programmes' components implemented, not implemented or modified. Average scores were calculated for major components, and presented as percentages.

Four major domains (22 components) were: organization & management, capacity-building, community work and surveillance. The full content of the tool is provided in the Appendix 3.17.

(Ryman et al., 2011) evaluated extent and quality of implementation of the reaching every district approach (RED), aimed to increase immunization coverage in 70 districts of Northern Sudan. The instrument contained 39 items, which were weighted by an expert opinion team. Each item was also assigned a level (in number, percentage or YES option) for achieving high RED implementations score (See Appendix 3.18, Table 27). The total score for each component was calculated by summing scores of indicators. The overall score was calculated by summing component scores. This overall score was further scaled on 10-point scale (10=fully implemented), and categorized into 'low' (<5.28), medium (5.28-6.31), high (>6.31) levels of implementation.

Tool components and indicators within each component were based on the five RED objectives set by WHO guidelines (available at: http://www.who.int/immunization_delivery/systems_policy/AFRO-RED-guide_2008.pdf) and included: outreach (3 items: e.g., % of planned different session), supervision (10 items: number of supervision visits, quality of supervision records), monitoring for action (11 items: availability of monitoring charts and monthly reports, number of review meetings), planning and management of resources (10 items: functioning refrigerators, vaccine stock out), and community links (5 items: percentage and types of social mobilization activities). An example of indicator description is provided in Appendix 3.18, Table 28. The full list of indicators, their weights and the level of for achieving high RED implementation score is available in Appendix 3.18.

(Sogarwal and Bachani, 2011) provided process evaluation of implementation of 197 community care centres for people living with HIV in 23 states and 158 districts in India to determine continuity of the scheme. The instrument for community centres implementation assessment is 0-100 percentage scale where each of 18 indicators under six domains is assigned a weight. The indicators are weighted "on the basis of predetermined options". However, how and on which basis these options were determined is not clear from the article. The scores were summed. To produce final summary scores, two semi-structured tools exploring barriers and problems for optimal functioning of the centres were used in addition to a quantitative scale (again, it is not clear how they were added to produce a final score). The final scores were ranged into four groups: "A" (80%+), "B" (70%-79%), "C" (60%-69%), "D" (<60%). These scores were interpreted as A=contract continued, B=contract continued with recommendations, C=temporary extension with major recommendations, D=contract cancelled.

The content of the tool includes six domains: physical infrastructure, systems, human resources, services, financial management, monitoring, and evaluation. The instrument design was based on the operational guidelines of the community care centres. The full content of the tool is available in the Appendix 3.19.

The content of the questionnaire was developed using the project protocol, "Action plan for the TIPPA Project 2000-2003".

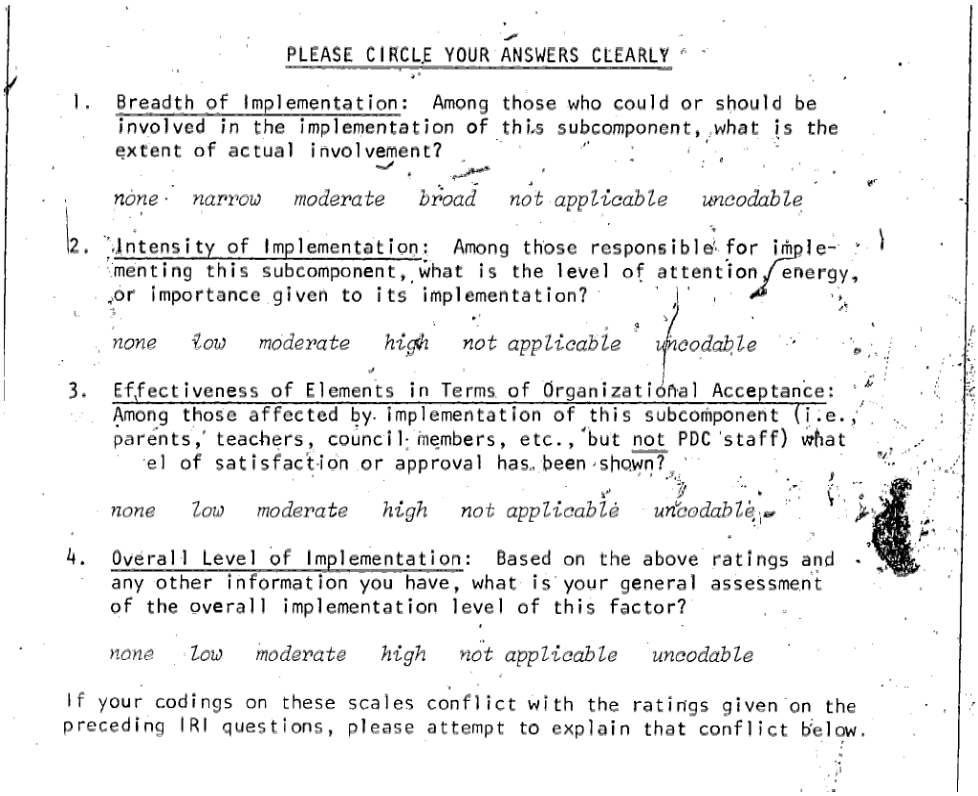
(Yumo et al., 2011) assessed implementation of the collaborative TB/HIV activities in a rural district hospital in Cameroon. The implementation level was set as a score, "a rate calculated using as numerator the total number of recommended activities effectively implemented in the hospital and as denominator the total number of recommended activities for the operational level". The level was presented in percentages. For example, if 4 of the 8 recommended activities were implemented, the implementation score was 50%.

The criteria to assess the implementation was based on the WHO recommended TB/HIV activities and indicators in "A Guide to monitoring and evaluation for collaborative TB/HIV activities: and included: HIV counselling and testing for TB patients; HIV prevention among TB patients; preventive therapy to TB/HIV co-infected patients; HIV care and support to TB/HIV patients; antiretroviral therapy to TB/HIV coinfected patients; TB screening and diagnosis in HIV positive patients; preventive therapy for HIV patient with latent TB; TB prevention control in the hospital" (the full list of indicators is available at http://www.panafrican-medjournal.com/content/article/10/30/material/10-30_1.pdf

Appendix 4. Studies' tools samples.

Appendix 4.1. (Smith, 1977)

Table 8. An example of judgemental rating scales for the subcomponent: Development of a diagnostic and evaluative system and individualised programme for children.



PLEASE CIRCLE YOUR ANSWERS CLEARLY

1. Breadth of Implementation: Among those who could or should be involved in the implementation of this subcomponent, what is the extent of actual involvement?

   none narrow moderate broad not applicable uncodable

2. Intensity of Implementation: Among those responsible for implementing this subcomponent, what is the level of attention, energy, or importance given to its implementation?

   none low moderate high not applicable uncodable

3. Effectiveness of Elements in Terms of Organizational Acceptance: Among those affected by implementation of this subcomponent (i.e., parents, teachers, council members, etc., but not PDC staff) what el of satisfaction or approval has been shown?

   none low moderate high not applicable uncodable

4. Overall Level of Implementation: Based on the above ratings and any other information you have, what is your general assessment of the overall implementation level of this factor?

   none low moderate high not applicable uncodable

If your codings on these scales conflict with the ratings given on the preceding IRI questions, please attempt to explain that conflict below.

Appendix 4.2. (Rubin et al., 1982)

Table 9. Sample from the field-based implementation rating scale.

Activity or Attribute Associated with the Programme	Criterion level	Monitoring/Documenting Activity and Data Source	Level of Implementation	Criterion level met or not met
1.Visiting homes of programme children by paraprofessionals	50% of paraprofessional's employment time At least 75% of the scheduled home visits will be completed for at least 80% of the programme children	Check for the reported percentage of time in the paraprofessional's weekly reports of the randomly selected paraprofessionals. Observe paraprofessionals in a sub-sample of their visits for a week. Do this for at least five randomly selected paraprofessionals	50% Home visits 50% Classroom 75%	Yes Yes
2.Spending time in institutional by the paraprofessional in the classroom	At least 50% of paraprofessional's employment time	Using the taxonomy of Classroom Activities Checklist, observe in the classroom of the paraprofessionals who were randomly selected	45%	No
3.Attending Policy Advisory Committee meetings by parents	At least 35% of the parents attend one meeting during the school year	Check for randomly selected sign-in attendance sheets for those parents affiliated with the randomly selected paraprofessionals Observe a meeting and verify the sign-in sheets at the meeting	40%	Yes

Appendix 4.3. (Glanz et al., 1992)

Table 10. Essential components of the physician-based nutrition programme.

Category	Components
Screening	1. Measurement of blood cholesterol for all patients age 20 or older, at least every 5 years 2. Identification of other cardiovascular risk factors (e.g., smoking, hypertension) 3. Assignment to receive treatment if warranted 4. Physician message: communication of screening results to (i) inform patient of cholesterol value and risk level, and (ii) emphasize importance of dietary change and recommended counselling
Cholesterol counselling	5. Assessment of eating pattern 6. Informing patient of cholesterol value and risk level 7. Discussion of eating pattern assessment results 8. Instruction on cholesterol-lowering diet, including food- and brand-specific recommendations 9. Discussion of and setting specific eating behaviour goals 10. Monitoring cholesterol and eating pattern changes
Tracking and monitoring: office-wide activities to create clinic cholesterol management system	11. Tracking patient appointments and follow-up visits 12. Tracking billing and reimbursement for patient counselling 13. Recording and summarizing CMP progress 14. Creating communication networks to keep physicians and office staff informed of CMP progress

Appendix 4.4. (Gold et al., 1993)

Table 11. Indices of implementation of Medicaid expansions, their components and items within components, and their percentage contribution to the overall index of strength of implementation.

Indices and components	%
1. Index of Eligibility and Enrolment	**50.00**
Eligibility component	***25.00***
of eligibility expansion	5.00
Timing of eligibility changes	5.00
Total amount of change in eligibility levels	5.00
Assets test dropped--timing	5.00
Continuous eligibility--timing	5.00
Enrolment component	***20.00***
Presumptive eligibility	3.64
Outstationing	3.64
Expedited review procedure	1.82
Availability of applications at different types of sites	1.82
Availability of applications at four or more types of sites	1.82
Applicants permitted to mail in forms	1.82
Mailed-in application processed without interview	1.82
Shortened application forms	1.82
Health care providers trained to assist women with applications	1.82
Geographic coverage of presumptive eligibility and outstationing component	***5.00***
% of poor women aged 15-44 living in a county with a presumptive eligibility site	2.50
% of poor women aged 15-44 living in a county with an outstationing site	2.50
2. Index of Enhanced Services	**15.00**
Medicaid enhanced services component	***10.00***
Care coordination/case management	1.43
Psychosocial risk assessment	1.43
Nutrition counselling	1.43
Health education	1.43
Psychosocial counselling	1.43
Home visiting	1.43
Transportation	1.43
Availability component	***5.00***
No. of enhanced services provided by at least half of MCH-funded prenatal care sites (7 services, each 0.714)	5.00

3. Index of Outreach Activities	**20.00**
Publicity component	***10.00***
Media campaign	0.62
Earliness of initiation of media campaign	1.25
Total no. of types of media strategies used to publicize expansions (6 types, each 0.626[1])	3.75
Total no. of types of sites where information was posted (7 types, each 0.626[2])	4.38
Hot-line component	***10.00***
Hot-line capacity	3.20
Toll-free hot line	0.36
Earliness of initiation of hot line	0.71
No. of calls to hot line as % of Medicaid-funded deliveries	0.71
No. of hours per day hot line is operational	0.71
Hot line operational on evenings or weekends	0.71
Types of information given by hot line	3.96
Location of prenatal care providers	0.36
Providers' hours of operation	0.36
Cost of prenatal care	0.36
Availability of subsidized prenatal care	0.36
Availability of public transportation to or from prenatal care sites	0.36
Location of presumptive eligibility sites	0.36
Location of outstationing sites	0.36
Location of social services agencies	0.36
Any other information	0.36
Scheduling of prenatal care appointments by hot-line operators	0.71
Hot-line publicity	
Total no. of types of media strategies used to publicize hot line (8 types, each 0.357[3])	2.86
Index of Reimbursement Generosity	**15.00**
Change in global OB fee, 1986-1991 (in 1986 dollars)	7.50
1991 level of reimbursement, standardized for state variation in medical costs	7.50
Overall Index of Strength of Implementation	**100.00**

Note: Percentages may not add to totals because of rounding.

Appendix 4.5. (Teague et al., 1998)

Table 12. Programme criteria for fidelity to assertive community care.

HUMAN RESOURCES STRUCTURE/ COMPOSITION H1 Small Caseload client/provider ratio of 101. H2 Team Approach provider group functions as team rather than individual practitioners; clinicians know & work with all clients H3 Program Meeting, program meets frequently to plan, review services for each client H4. Practicing Team Leader: supervisor of front-line clinicians provides direct services. H5. Continuity of Staffing' program maintains same staffing over time. H6 Staff Capacity: program operates at full staffing. H7 Psychiatrist on Staff, at least one full-time psychiatrist per 100 clients assigned to program H8 Nurse on Staff >2 full-time nurses per 100 clients. H9 Substance Abuse Specialist on Staff: >2 staff with 1 yr training/clinical exp. in substance abuse treatment H10 Vocational Specialist on Staff. >1 staff member with >1 yr training/exp in vocational rehab/support. H11 Program Size: sufficient absolute size to provide consistently the necessary staff diversity & coverage **ORGANIZATIONAL BOUNDARIES** 01 Explicit Admission Criteria: clearly identified mission to serve particular population, measurable, operationally defined criteria to screen out inappropriate referrals. 02 Intake Rate, takes clients in at a low rate to maintain a stable service environment 03 Full Responsibility for Treatment Services, as well as case management/psychiatric services, program directly provides counselling/psychotherapy, housing support, substance abuse, employment, S rehab services 04 Responsibility for Crisis Services 24-hour coverage of psychiatric cases	05 Responsibility for Hospital Admissions program is involved in hospital admissions 06. Responsibility for Hospital Discharge Planning program is involved in planning hospital discharges 07 Time-Unlimited Services, program closes no cases, remains point-of-contact for all clients as needed. **NATURE OF SERVICES** S1 In-Vivo Services program monitors status, develops community living skills in community rather than office S2 No Dropout Policy, program engages/ retains clients at mutually satisfactory level S3. Assertive Engagement Mechanisms' uses street outreach, plus legal mechanisms (e g , representative payees, probation/parole, OP commitment) as indicated. S4 Intensity of Service high total amount of service time, as needed. S5 Frequency of Contact high number of service contacts, as needed. S6 Work With Support System with or without client present, program provides support/skills for client's support network: family, landlords, employers S7 Individualized Substance Abuse Treatment >1 program member provides direct treatment & substance abuse treatment for clients w/substance use disorders S8. Dual Disorder Treatment Groups group modalities used as tx strategy for people w/substance disorders S9. Dual Disorders Model: uses a stage-wise treatment model that is nonconfrontational, follows behavioural principles, considers interactions of mental illness & substance abuse, S has gradual expectations of abstinence S10. Role of Consumers on Treatment Team: clients involved as team members providing direct services.

Table 13.Sample of structure and content of the ACT Scale.

Table 2

ANCHORS FOR THREE SAMPLE ITEMS

	SCALE				
ITEM	1	2	3	4	5
H3. Program Meeting. Program meets frequently to plan & review services for each client	Service planning for each client usually occurs once a month or less frequently	At least twice a month but less often than once a week	At least once a week but less often than twice a week	At least twice a week but less often than 4 times weekly	At least 4 days a week & reviews client each time, even if briefly
O4. Responsibility for Crisis Services. Program has 24-hr responsibility for psychiatric crises	No responsibility for handling crises after hours	Emergency service has program-generated protocol for clients	Program available by phone, largely in consulting role	Provides emergency service backup	Provides 24-hour coverage
S1 In-Vivo Services. Program works to monitor status and develop living skills in community rather than in office	Less than 20% time in community	20%–39%	40%–59%	60%–79%	80% of total service time in community

Appendix 4.6 (Orwin et al., 2000)

Table 14. Formula for homeless programme implementation scales.

(1) Total services 5 R *(MIN[nQi,TIP])/nStnQt,* where *nQi* = number of quarters that the participant received services for service *i*, *TIP* is the total time-in program for that participant, *nSt* = number of total possible services (the number of services on the QRF), and *nQt* is the theoretical maximum of the intervention's duration in quarters. *MIN[nQi,TIP]* takes the number of person-quarters that the participant was receiving services for service *i or* the total *TIP*, whichever is shorter [the rationale is that logically, the program cannot be providing the participant services for longer than he or she is in the program (experience working with the QRF data suggested that when discrepancies occurred, the start and exit dates were reported more accurately than numbers of service quarters)]. *nSt* is a constant of 48 (the number of service categories on the QRF), and *nQt* is a constant of 6. Six was used as the intended duration constant because 18 months (6 quarters) was the maximum intended duration of any Cooperative Agreement intervention; this forced all scores to range between 0 and 1.0, and provided a "level playing field" for comparing intervention groups across sites (an *adjusted* total-services scale was also derived specifically for within-site comparisons. The formula does not change; however the value of *nQt* was set at the intervention's own intended duration, rather than that of the longest Cooperative Agreement intervention. So, for example, if the intended duration of a given intervention is 4 months, an individual's score on the adjusted total-services scale would be 18/4 (or 4.5) times the value of the unadjusted score. The adjusted version can then range from 0 to 1.0 for each intervention, while the unadjusted version can only reach 1.0 for the longest (18 months) interventions). Modifications to the formula were occasionally required, for example, if a project's intervention length was indefinite and did not have exit dates.
2) *Fidelity = R (MIN[nQi(f),TIP])/nSt(f)nQt,* where nQi(f) = number of quarters that the participant received services for "featured" service i(f), TIP is (as before) the total time-in-program for that participant, nSt(f) = number of total featured services provided by that intervention (the number of "featured" services on the QRF), and nQt is (as before) the theoretical maximum of the intervention's duration in quarters. MIN [nQi(f),TIP] takes the number of person-quarters that the participant was receiving services for featured service i(f) or the total TIP, whichever is shorter. Like the total-services scale, the fidelity scale ranges between 0 and a theoretical maximum of 1.0. As with the total services scale, modifications to the formula were occasionally required, for example, if a project's intervention length was indefinite and did not have exit dates.
Leakage = R (MIN[nQi(l),TIP])/nSt(f)nQt where nQi(l) = number of quarters that the participant received services for service i(l) that were "leaked" services, that is, services not intended to be available for that group but available or featured in another group, TIP is (as before) the total time-in-program for that participant, nSt(l) = number of total leaked services potentially possible for that intervention, and nQt is (as before) the theoretical maximum of the intervention's duration in quarters. MIN[nQi(t),TIP] functions as in the other two scales. Like the prior scales, the leakage scale ranges between 0 and a theoretical maximum of 1.0.

Appendix 4.7. (Vinson et al., 2001)

Table 15.Domains, their attributes and definitions for evaluation of implementation fidelity of the system-of-care model.

INFRASTRUCTURE	
1.Goals and Vision	The aims and purposes 1) reflect a consensus among participants about optimal provision of services 2)are manifested in systems strategies, policies and procedures 3)are used to guide decisions; 4)are consistently implemented
2.Target Population	Participants' shared definition of the children and families served by the system of care, the extent of the system of care's ability to serve consumers in all catchment areas, and the diversity of referral sources
3.Interagency structure	A consortium of core agency, provider, family, and community representatives with governance, programme planning, and budgetary authority. The structure operates in a consistent manner, and participant roles and functions are defined an institutionalized in formal interagency agreements.
4.Community based	Local determination of the organization and array of services whereby the system can serve children locally, thus reducing inappropriately restrictive, out-of-community care. Services are conveniently located throughout the catchment area, focus on local children and their families, and reflect input from community representatives.
5.Human resources development	Full staffing of the system with qualified individuals who have been effectively trained in system-of-care principles. Family input is sought with regard to staff hiring, training, and performance issues
6.Case review	Cross-agency deliberation regarding the service array and service options that can and should be provided to individual children and families. A variety of participants are involved so that support for decisions is generated throughout the system.
7.Evaluation	Assessment of outcomes and service quality; results actively guide decision making, and the system of care can be held accountable for the goals it sets.
8.Funding	The system's financial capacity to achieve its goals using a local, cross-agency infrastructure for fiscal management.
9.Communication	The effective, efficient, systematic dissemination of information among and between participants, accomplished through various informal and formal mechanisms.
10.Service Array	The diversity of mental health and non-mental health services offered: growth in the number and types of services available, and expansion of the system's service capacity.
SERVICE DELIVERY	
Case management	Processes and tasks that facilitate a match between individual needs and the types and intensities of services available. Case managers should a)be given responsibility for a manageable number and mix of cases; b)coordinate services; c)monitor progress; d)facilitate communication across providers; e)help families to identify their needs, strengths, and available resources; f)help families plan services; g)listen to and respect family input; and h)vary the intensity of service coordination to meet child and family needs. Case managers' decisions should be accepted across agencies so that the system operates effectively

Service delivery coordination	Cross-agency coordination of a child and family's services as achieved through a)agreed-upon eligibility criteria and shared intake processes across core agencies; b)systematic information sharing; c)routine updates and recording of all services received; d)the use of shared service plans across agencies; e)procedures to assure continuity of care and smooth transitions.
Service accessibility	Physical, geographic, linguistic, and financial services accessibility. Outreach is regular and effective, entry points into the system are clearly identified, waiting lists are shortened or eliminated, and no family or refused services as a result of system capacity issues.
Family focus	The family, as broadly conceived, is the service provision unit. Families are involved in decision making and feel respected at all levels of the system, formal family involvement policies are enacted, family feedback is actively sought to evaluate the system and hold it accountable, and family advocacy and empowerment are promoted.
Individualised service planning	Services are identified for and by families and youth, services are tailored to reflect unique strengths and needs, and the service plan incorporates input from families, youth, multiple agencies, and cross-disciplinary service providers. Service plan content is detailed, with specific, achievable, strength-based behavioural and treatment goals.
Cultural competence	The system of care's sensitivity and responsiveness to the culture, gender, race, ethnicity, religion, national origin, sexual orientation, physical disability, educational, and socio-economic status of families. This is formalized in policies, procedures, outreach and advocacy efforts, training, service array, the service delivery framework, and the recognition of the importance of existing community support networks. Staff and service providers are knowledgeable, skilled, and aware of cultural issues in the community.

Appendix 4.8. (Paulsell et al., 2002)

Table 16. Program elements included in the early head start implementation rating scales.

Early childhood development and health services

- Frequency of child development services
- Developmental assessments
- Follow-up services for children with disabilities
- Child health services
- Child care services
- Parent involvement in child development services
- Individualisation of child development services
- Group socialisation activities

Family and community partnerships

- Family partnerships Individualised family partnership agreements
- Availability of services
- Frequency of family development services
- Parent involvement
- Community partnerships Collaborative relationships with other service providers
- Advisory committees
- Transition plans

Management systems and procedures

- Staff development Supervision
- Training
- Staff retention
- Compensation
- Staff morale
- Program management Policy council
- Communication systems
- Goals, objectives, and plans
- Self-assessment
- Community needs assessment

Appendix 4.9. (Cooley et al., 2003)

Table 17. Medical home index domains and themes.

Domains	Themes
1. Organizational capacity	1.1 Mission of the practice
	1.2 Communication/access
	1.3 Access to medical records
	1.4 Office environment
	1.5 Family feedback
	1.6 Cultural competence
	1.7 Staff education
2. Chronic condition management	2.1 Identification of CSHCN
	2.2 Care continuity
	2.3 Continuity across settings
	2.4 Cooperative management with specialists
	2.5 Supporting transition to adult services
	2.6 Family support
3. Care coordination	3.1 Role definition
	3.2 Family involvement
	3.3 Child and family education
	3.4 Assessment of needs/plans of care
	3.5 Resource information and referrals
	3.6 Advocacy
4. Community outreach	4.1 Community assessment of needs of CSHCN
	4.2 Community outreach to agencies and schools
5. Data management	5.1 Electronic data support
	5.2 Data retrieval capacity
6. Quality improvement	6.1 Quality standards (structures)
	6.2 Quality activities (processes)

Table 18. Example of the scale for the Care Coordination domain.

Domain 3 Care Coordination

THEME:	*Level 1*	*Level 2*	*Level 3*	*Level 4*
#3.4 Assessment of Needs/ Plans of Care	Presentation of CSHCN with acute problems determines how needs are addressed. PARTIAL COMPLETE	*PCPs* identify specific needs of CSHCN; follow-up tasks are arranged for, or are assigned to families &/or available staff. PARTIAL COMPLETE	The child with special needs, family, and *PCP* review current child health status and anticipated problems or needs; they create/revise action plans and allocate responsibilities at least 2 times per year or at individualized intervals. PARTIAL COMPLETE	In addition to Level 3, the PCP/staff and families create a written plan of care that is monitored at every visit; the office *care coordinator* is available to the child and family to implement, update and evaluate the care plan. PARTIAL COMPLETE

Appendix 4.10. (Hacker & Washington, 2004).

Table 19. Scale to evaluate large-scale organisational change efforts.

1. Well defined Key Result Areas and Goals- ❑ Linked to the Ministry Vision. ❑ Goals defined as appropriate. ❑ Cross Ministry Key Result Areas identified as needed. ❑ Measurable. ❑ Targets.	
2. Well defined Objectives- ❑ Linked to specific Key Result Areas. ❑ Strategies defined as needed. ❑ Prioritized critical few. ❑ Measurable. ❑ Longer term targets. ❑ Annual Targets. ❑ Annual Performance Plans include actions to achieve the Objectives.	
3. Well Defined Measurement Process- ❑ Accepted measures. ❑ Understanding of measures. ❑ Reliable data sources. ❑ Reliable survey mechanics. ❑ Plotted as time-series. ❑ Control charted. ❑ Special causes investigated. ❑ Control limits revised as appropriate. ❑ Data interpretations accurate. ❑ Random variation understood.	
4. Well Established Reviews- ❑ Quarterly reviews conducted with PS. ❑ Monthly reviews conducted with Department Heads. ❑ Status reports generated. ❑ Follow up from prior reviews. ❑ Lessons learned discussed. ❑ Discussion includes plans for next quarter and longer. ❑ Performance accountability exists.	
5. PMS Responsibilities Defined and Visible[1] ❑ Permanent Secretary. ❑ Department Heads. ❑ WITs. ❑ PICs. ❑ BNPC Consultants. ❑ DMPSM	
6. Continuous Improvement Processes Exist- ❑ PMS process annually assessed. ❑ WITs' efforts aligned with Key Result Areas. ❑ Training in PMS provided systematically and broadly. ❑ New employees trained in PMS. ❑ Systematic development of PICs' PMS skill level. ❑ Communication plans keep organization aware of the status of PMS. ❑ Internal PMS benchmarking across Ministries.	

[1] These are specific organizations / roles / individuals involved in the PMS effort in Botswana. These names would be changed to reflect the key personnel in the specific's organization's change efforts.

Appendix 4.11. (Hebert & Veil, 2004).

Table 20.List of indicators used to rate the implementation of the ISD system.

Indicators	Rating points
Coordination	20
1. Presence of a structure designed to enhance cooperation between partners	3
2. All partners concerned represented	3
3. Representatives stability over time	3
4. Representatives participate regularly	3
5. Partners informed of how services are changing (or not changing)	4
6. Partners criticize the organization of the services change process	4
Single entry point	20
1. Presence of a single entry point in each local area	5
2. Clearing functions done by dedicated professionals	5
3. Professionals use a validated screening instrument to identify eligible frail elderly	5
4. Follow-up with older people in the group at high risk of functional decline	5
Case management 20	20
1. Variation between actual number of case managers and number needed according to proportion of senior	10
citizens in the area	10
2. Variation (above or below) between actual average caseload and recommended caseload	
Single assessment tool and case-mix classification	15
1. Percent of clients under case management evaluated with SMAF tool	5
2. Percent of partners systematically using SMAF tool with their elderly patients	5
3a. Use of the case-mix classification system (ISO-SMAF profiles) for efficient utilization of resources	5
3b. Use of the ISO-SMAF profiles system as a new standard for financing services	
Computerized clinical chart	15
1. Availability of a computer program for sharing clinical information in real time	5
2. Sufficient number of computers for all partners	5
3. Utilization of the computerized computer chart by partners	5
Individualized service plan	10
1. Percent utilization of the individualized service plan by case managers (as indicated in the clinical files of patients under case management)	
TOTAL	100

Appendix 4.12. (Kansaharo et al., 2005)

Table 21. Statements or action points to measure the implementation rate of the TIPPA project .

Actions/Statements	Very well to quite well implemented (%)	Implemented to some extent	Not implemented or do not know (%)
1) Information sources concerning drug information are easily available in our pharmacy 2) Our pharmacy provides complete privacy for patient counselling 3) Our pharmacy arranges regular in-house training for patient counselling 4) Our pharmacy has made sure that the process of writing prescriptions with the computers is now more effective to leave more time for patient counselling 5) The general contracts and action plans concerning drug information on prescription drugs have been drawn up in our pharmacy 6) The general contracts and action plans concerning drug information on 7) OTC-drugs have been drawn up in our pharmacy 8) Our pharmacy's education plan is made once a year 9) Our pharmacy co-operates in patient counselling with other local health care professionals 10) Our pharmaceutical staff has the possibility to take part in developing patient counselling 11) Our pharmacy has a counselling development plan 12) Development discussions concerning patient counselling are held regularly in our pharmacy 13) Our pharmaceutical staff has their own personal education plan 14) Our pharmacy has made sure that our staff has the ability to use the Internet as a tool in patient counselling in our pharmacy 15) Feedback from the customers concerning counselling will be recorded and discussed among the staff 16) Our pharmacist is authorized to be responsible for developing drug information about OTC-drugs 17) Our pharmacist is authorized to be responsible for developing drug information about the prescription drugs			

Appendix 4.13. (Pearson et al., 2005)

Table 22.Implementation intensity components.

1. **Delivery system redesign**
    a. Care management roles
    b. Team practice
    c. Care delivery/coordination
    d. Proactive follow-up
    e. Planned visit
    f. Visit system change
2. **Self-management support strategies**
    a. Patient education
    b. Patient activation/psychosocial support
    c. Self-management assessment
    d. Self-management resources and tools
    e. Collaborative decision making with pts.
    f. Guidelines available to patients
3. **Decision support**
    a. Guideline institutionalization and prompts
    b. Provider education
    c. Expert consultation support
4. **Information support**
    a. Patient registry system
    b. Use of information for care management
    c. Feedback of performance data
5. **Community linkages**
    a. For patients
    b. For community
6. **Health systems support**
    a. Leadership support
    b. Provider participation
    c. Coherent system improvement and spread
    d. Chronic Care Model overall

Appendix 4.14. (Grizzard et al., 2006)

Table 23. Nurse manager survey of ten steps implementation.

Step 1: Have a written breastfeeding policy that is routinely communicated to all healthcare staff **Have a hospital-wide policy that affects breastfeeding care plan(s)**	
1. Does your hospital have a written breastfeeding policy or protocol?	0-no 1-yes
2. What type of policy is it?	1-institutional policy concerning breastfeeding 2-nursing protocol concerning breastfeeding 3-other (please describe fully) 4-uncertain
3. Does the policy state that a physician's order is required in order for breastfed infants to receive anything but breast milk?	0-no 1-yes 2-uncertain
4. If yes to 3, can this policy be overridden by a parent's request?	0-no 1-yes 2-uncertain
5. Is this breastfeeding policy visibly posted on the postpartum floor?	0-no 1-yes 2-uncertain
6. Is this breastfeeding policy made readily available to postpartum staff members?	0-no 1-yes 2-uncertain
7. Are non-nursing staff members like nursery aides or receptionists routinely made aware of the breastfeeding policy?	0-no 1-yes 2-uncertain 3-n/a
Step 2: Train all health care staff in skills necessary to implement this policy	
8. Is there some type of "breastfeeding orientation" for new personnel on your postpartum staff?	0-no 1-yes
9. Who attends this orientation?	1-physicians and nurses 2-nurses only 3-physicians only 4-non-nursing postpartum staff members only 5-physicians, nurses, and non-nursing postpartum staff members. 6-non-nursing staff and nurses
10. Are there opportunities for breastfeeding education (apart from breastfeeding orientation) available for postpartum staff at your hospital?	0-no 1-yes
11. What type of breastfeeding education (other than breastfeeding orientation) can they receive AT YOUR HOSPITAL?	1-informal lectures offered by hospital 2-written or audio-visual information 3-workshops on breastfeeding topics 4-other
12. How many total hours of breastfeeding education (other than breastfeeding orientation) are typically offered per year?	1-less than 5 hours 2-5-10 hours 3-11-16 hours 4-17-20 hours 5-more than 20 hours per year 6-uncertain
13. Who on your postpartum staff utilizes these breastfeeding educational offerings?	1-physicians and nurses 2-nurses only 3-physicians only 4-non-nursing postpartum staff members only, i.e. clinical aides, secretaries 5-physicians, nurses, and non-nursing postpartum staff members 6-non-nursing staff and nurses
14. Do you encourage postpartum staff members to receive breastfeeding education OUTSIDE OF YOUR HOSPITAL?	0-no 1-yes
Step 3: Inform all pregnant women about the benefits and management of breastfeeding **Prenatal education classes emphasize & support breastfeeding**	
15. Does your hospital offer childbirth education or preparation classes for parents?	0-no 1-yes

	2-don't know/uncertain
16. What percentage of total class time in these childbirth education classes is spent discussing breastfeeding and its benefits?	1-more than 90% 2-75-90% 3-25-74% 4-10-24% 5-less than 10% 6-uncertain/do not know
17. Is breastfeeding routinely presented as the standard way to feed an infant in these classes?	0-no 1-yes 2-don't know/uncertain
18. Does your hospital offer specific prenatal breastfeeding classes?	0-no 1-yes 2-don't know/uncertain
19. Are audio-visual and/or printed materials about breastfeeding available for parents on the postpartum floors?	0-no 1-yes
Step 4: Help mothers initiate breastfeeding within half an hour of birth	
20. Are healthy term infants routinely required or do they routinely required to go to the Newborn Nursery after delivery?	0-no 1-yes
21. If yes to 20, how many hours after delivery are healthy term infants routinely sent to the Newborn Nursery?	1-immediately after delivery 2-less than ½ hour 3-1/2 to 1 hour 4- 1 to 1, ½ hour 5-2 hours or more after delivery
22. If yes to 20, how long do healthy term infants routinely remain in the Newborn Nursery during the initial nursery stay after delivery?	1-less than ½ hour 2- 1/2 to 1 hour 3-1 to 1, ½ hours 4- 2 -2,1/2 hours 5-3 hours or more
23. How soon after delivery are healthy term infants routinely put to the breast, if their mothers have indicated a desire to breast-feed?	1-immediately following delivery 2-less than 1/2 hour 3-1 to 1, ½ hours 4- 2 to 2, ½ hours 5-2 hours or more
Step 5: Show mothers how to breast-feed and how to maintain lactation even if they should be separated from their infants	
24. Do you have at least one IBCLC on-staff as a permanent staff member whose primary responsibility is lactation support?	0-no 1-yes
25. If "Yes" to 24, how many dedicated IBCLCs are employed as FULL-TIME equivalents by your hospital?	0-zero 1-one 2-two 3-three 4-four or more
26. How many dedicated IBCLCs are employed per diem by your hospital?	0-zero 1-one 2-two 3-three 4-four or more
27. If "No" to 24, do you have anyone else as a permanent staff member who serves as a dedicated lactation consultant?	0-no 1-yes
28. Is there an IBCLC or LC "on-call" for consults or questions outside of 8-5 business hours?	0-no 1-yes
29. Do postpartum staff members routinely give mothers oral breastfeeding instructions?	0-no 1-yes
30. Does your nursing staff routinely demonstrate manual breast milk expression to its breastfeeding mothers?	0-no 1-yes 2-uncertain
31. How often does your nursing staff routinely assist mothers who are separated from their infants for medical reasons with establishing lactation?	1-occasionally 2-some of the time 3-most of the time 4-always 5-never
32. How many hours after delivery are mothers who are separated from their infants for medical reasons routinely offered an electric breast pump?	1-less than 6 hours after delivery 2-7-12 hours 3-12-23 hours 4-24 hours or more after delivery 5-not routinely provided with an

	electric breast pump following delivery 6-N/A
Step 6: Give newborn infants no food or drink other than breast milk unless medically indicated Infants are not routinely supplemented with non-breast milk nutrition	
33. Does your hospital routinely give breast milk as the first feeding to infants whose mothers have indicated that they wish to breastfeed?	0-no 1-yes
34. How often are breastfed infants on your postpartum floor routinely given non-breast milk nutrition, e.g., supplemented with formula in the nursery overnight?	1-occasionally 2-some of the time 3-most of the time 4-always 5-never
35. Does your hospital receive free samples of formula from formula companies?	0-no 1-yes
36. Does your staff routinely give the same commercial discharge packet to both breastfeeding and non-breastfeeding mothers?	0-no 1-yes
37. If no to 36, does the breastfeeding packet include formula samples?	0-no 1-yes 2-n/a 3-uncertain
38. Does your staff routinely give formula samples (other than those that might be found in the commercial discharge packet) to breastfeeding mothers?	0-no 1-yes
39. Does your hospital display promotional materials (posters, pens, id tags, calendars, notepads etc.) for formula ANYWHERE on your postpartum or labour floors?	0-no 1-yes
Step 7: Practice rooming-in and allow mothers and infants to remain together 24 hours a day Promote rooming-in	
40. Is 24-hour "rooming-in" available at your hospital?	0-no 1-yes
41. Does your staff routinely encourage mothers to "room-in" with their babies 24 hours a day?	0-no 1-yes
42. How often do breastfeeding mothers on your hospital's postpartum floor typically room-in during the day and send their babies to the nursery at night?	1-occasionally 2-some of the time 3-most of the time 4-always 5-never
43. Do mothers have to meet specific requirements (i.e., must be ambulatory) set down by a protocol orvpolicy before they are allowed to room-in?	0-no 1-yes
Step 8: Encourage feeding on demand Moms are instructed about how to feed on demand	
44. Are mothers routinely or as part of a nursing care plan instructed about feeding cues in order to encourage them to feed their babies on demand?	0-no 1-yes
45. How often are breastfeeding mothers routinely taught how to assess efficacy of feeding sessions(e.g. checking for latch & swallowing)?	1-occasionally 2-some of the time 3-most of the time 4-always 5-never
46. How often are breastfeeding mothers routinely offered the opportunity to feed their infants at night if so desired?	1-occasionally 2-some of the time 3-most of the time 4-always 5-never
47. Are visiting hours limited at your hospital?	0-no 1-yes
48. If yes to 47, how many hours a day are visitors permitted to visit?	1-2-5 hours per day 2-6-8 hours per day 3-9-11 hours per day 4-12 hours a day or more
49. Are these visiting hours routinely enforced?	0-no 1-yes
Step 9: Give no additional teats or pacifiers to breastfeeding infants	
50. Are pacifiers available on your hospital postpartum floor for use by staff or patients?	0-no 1-yes
51. Are pacifiers routinely given on your hospital's postpartum floor?	0-no 1-yes
52. How often are pacifiers given without parents' request, if ever?	1-occasionally 2-some of the time 3-most of the time 4-always 5-never
53. Do postpartum staff members routinely explain reasons for limiting use of pacifiers in breastfed infants?	0-no 1-yes 2-uncertain

Step 10: Foster the establishment of breastfeeding support groups and refer mothers to them on discharge from the hospital or clinic	
54. Does your staff routinely refer breastfed mothers to community breastfeeding support services and resources—i.e., provide phone numbers for local La Leche League meetings or breastfeeding support groups—at time of discharge?	0-no 1-yes 2-uncertain
55. Does your staff routinely provide a hospital phone number for questions for its breastfeeding moms?	0-no 1-yes 2-uncertain
56. Does your staff routinely follow-up post discharge with its breastfeeding moms via telephone?	0-no 1-yes

Appendix 4.15. (D'Abbs et al., 2008)

Table 24. Seven dimension of health service systems from System Assessment Tool.

1. Organisational influence refers to the use of organisational influence to create a culture, and support organisational systems and mechanisms to promote safe, high quality care.

2. **Links within the community** and with external (outside) services refers to the extent to which the primary health care service uses linkages with other community resources and service providers to promote health and enhance the effectiveness of the service. It is primarily through this component of the tool that the quality of community or population based programs is assessed.

3. **Self-Management Support** refers to health service systems that support development of awareness and skills for clients and families to take a major role in maintaining their health, managing health problems, and promoting a safe and healthy environment.

4. **Decision Support** refers to systems to ensure easy accessibility of evidence based information to inform decisions related to individual clinical care, and the promotion and protection of the health of the service population.

5. **Delivery System Design** refers to the extent to which the design of service infrastructure, work flow, staffing and other service delivery systems maximises the potential effectiveness of the health service.

Appendix 4.16. (Edward et al., 2011).

Table 25. Balanced score card indicators and domains.

Domain A: Patients and Community

- Overall patient satisfaction
- Patient perceptions of quality
- Written Village Health Council activities in community

Domain B: Staff

- Health worker satisfaction
- Salary payments current

Domain C: Capacity for Service Provision

- 6 Equipment functionality
- 7 Drug availability
- 8 Family Planning availability
- 9 Laboratory functionality (CHCs & DHs)
- 10 Meeting minimum staff guidelines
- 11 Provider knowledge
- 12 Staff received training in last year
- 13 HMIS use
- 14 Clinical guidelines
- 15 Infrastructure
- 16 Patient records
- 17 Monitoring of TB treatment

Domain D: Service Provision

- 18 Patient history and physical exam
- 19 Patient counselling
- 20 Proper sharps disposal
- 21 Average new outpatient visits per month (BHC>750 visits)
- 22 Time spent with patient (> 9 minutes)
- 23 Provision of antenatal care
- 24 Provision of delivery care

Domain E: Financial Systems

- 25 User fee guidelines
- 26 Exemptions for poor patients

Domain F: Overall Vision

- 27 Females as percentage of new outpatients
- 28 Outpatient visit concentration index
- 29 Patient satisfaction concentration index

Appendix 4.17. (Perez et al., 2005).

Table 26. The strategy components.

1. **Organisation & management**
   Presence of Community Working Groups leading the strategy
   AaCP staff within Community Working Groups
   Community resources identified
   External resource mobilization
   Link with primary school established
   Other actors involved
2. **Capacity-building**
   Diagnosis, group work and participation
   Surveillance of risks and behaviours
   Planning and communication strategy
   Participatory evaluation
   Self-organized workshops
3. **Community work**
   Risk mapping
   Problems assessment
   Action plan
   Actions executed
   Communication strategy
   Elaboration of communicational materials
   Monitoring and evaluation
   Community strengths assessment
4. **Surveillance**
   Identification of environmental risks
   Identification of domiciliary risks
   Identification of associated behaviours

Appendix 4.18. (Ryman et al., 2011)

Table 27. RED Implementation components, indicators, indicator weighting scheme and level needed to achieve high implementation score indicator Individual

Component Indicator	Individual indicator	Individual indicator weight	Level of achieving high RED implementation score
Outreach	% of planned outreach sessions held	0.33	≥ 90%
	% of planned mobile sessions held	0.33	≥ 90%
	% of planned fixed sessions held	0.33	≥ 95%
Supervision	% of planned supervision to fixed sites by locality officer	0.10	≥ 80%
	% of fixed sites receiving ≥1 supervisory visit during the year	0.10	≥ 90%
	No. of monthly locality supervision reports submitted to state	0.10	12
	No. of supervision visits the locality received from the state	0.10	≥ 5
	No. of supervision visits the locality received from the federal	0.10	≥ 3
	% of outreach teams visited by locality officer	0.10	≥ 10%
	% of mobile teams visited by locality officer	0.10	≥ 10%
	Records available for all supervisory visits	0.10	Yes
	Quality of supervision records (as per peer-review guidelines)	0.10	100%
	Quality supervision records (as per peer-review guidelines)	0.10	100%
Monitoring for action	Monitoring charts available for all locality & all fixed Sites	0.09	Yes
	Monitoring chart with no graphing errors	0.09	Yes
	Monitoring chart with no calculation errors	0.09	Yes
	% of fixed sites with defaulter tracking for the entire year	0.09	≥ 95%
	Defaulter tracking success rate (recoup)	0.09	>80%
	% of fixed sites sending complete monthly reports	0.09	100%
	% of fixed sites sending timely monthly reports	0.09	100%
	No. of review meeting with fixed sites planned	0.09	12
	% of review meetings with fixed sites held	0.09	≥ 75%
	No. of monthly state level review meeting locality officer attended	0.09	12
	All cold chain equipment monitored twice daily	0.09	Yes

	(except Friday)		
Planning & management of resources	Completeness of micro-plan (as per peer-review guidelines)	0.20	100%
	% of locality refrigerators functioning through the year	0.10	≥ 90%
	Two trainings completed by locality officer	0.10	Yes
	% of vaccinators fully trained	0.10	100%
	Generator in locality office functional throughout the year	0.10	Yes
	Vaccine stock-out	0.05	0
	Vaccine stock-out duration	0.05	0
	Syringe stock-out	0.05	0
	Syringe stock-out duration	0.05	0
	Tally sheet stock-out	0.05	0
	Tally sheet stock-out duration	0.05	0
	Vaccine card stock-out	0.05	0
	Vaccine card stock-out duration	0.05	0
Community links	No. of different types of social mobilization activities planned	0.20	≥ 6
	% of planned social mobilization activities implemented	0.20	≥ 100%
	Expanded Programme on Immunization Friends Society member involvement	0.20	Yes
	Women's Union member involvement	0.20	Yes
	No. of other community groups involved in Expanded Programme on Immunization	0.20	≥ 2

Table 28. An example of core indictors description for "supportive supervision" components from WHO guide: Implementing the Reaching Every District approach: A guide for district health management teams

CORE INDICATORS FOR "SUPPORTIVE SUPERVISION"

Adapt and use the following process indicator—taken from the RED Monitoring Tool (Annex 2)—to monitor "supportive supervision" across districts.

% of DISTRICTS conducting at least (***minimum number***) of supportive supervisory visits to each of their health facilities per (***quarter***).

This core indicator is intended to remind districts teams (and inform those at higher levels of the health system) of the importance of conducting regular supportive supervision at each health facility according to national guideline frequency.

In addition to determining whether or not supportive supervision visits are conducted at the minimum frequency, districts and those supervising them will also want to review the frequency of visits to health facilities above minimum standards, the quality of visits (sufficient time, observation of sessions, on-the-job-training, etc.), and written feedback and follow up. This can be tracked by districts (self-assessment) and through supportive supervision visits to districts from higher levels of the system.

Appendix 4.19. (Sogarwal et al., 2011)

Table 29.Sample of the summary scores for the process evaluation of community care centres.

Indicators	Weights (%)	Scores
1 Physical infrastructure	15	0.0
Accessibility	5	0.0
Physical ambience and hygiene	5	0.0
Adequacy of infrastructure	5	0.0
2 Systems	20	0.0
Adequate and functional equipments	5	0.0
Adequacy of supplies and waste management	5	0.0
Infection control measures being followed	5	0.0
Interlinkages, referrals, and coordination	5	0.0
3 Human resource	15	0.0
Adequacy of staff	10	0.0
Training of staff	5	0.0
4 Services	25	0.0
Services delivery	5	0.0
Referral, outreach, and other support services	5	0.0
Level of patient satisfaction	10	0.0
Bed utilization	5	0.0
5 Financial and inventory management	15	0.0
Maintenance of accounts	5	0.0
Procurement and supplies	5	0.0
Expenditure and audit	5	0.0
6 Monitoring and evaluation	10	0.0
Regularity in updates—monitoring and evaluation	5	0.0
Record keeping at CCC	5	0.0
Final score	100	0.0

IDEAS project

IDEAS (Informed Decisions for Actions) aims to improve the health and survival of mothers and babies through generating evidence to inform policy and practice. Working in Ethiopia, northeast Nigeria and the state of Uttar Pradesh in India, IDEAS uses measurement, learning and evaluation to find out what works, why and how in maternal and newborn health programmes.

IDEAS is funded by a grant from the Bill & Melinda Gates foundation to the London School of Hygiene & Tropical Medicine.

ideas.lshtm.ac.uk

London School of Hygiene & Tropical Medicine

The London School of Hygiene & Tropical Medicine is a world-leading centre for research and postgraduate education in public and global health, with 4000 students and more than 1300 staff working in over 100 countries. The School is one of the highest-rated research institutions in the UK, and was recently cited as one of the world's top universities for collaborative research.

www.lshtm.ac.uk

London School of Hygiene & Tropical Medicine
Keppel Street, London, WC1E, 7HT, UK

t +44 (0)207 927 2871/2257/2317
w ideas.lshtm.ac.uk
@LSHTM_IDEAS