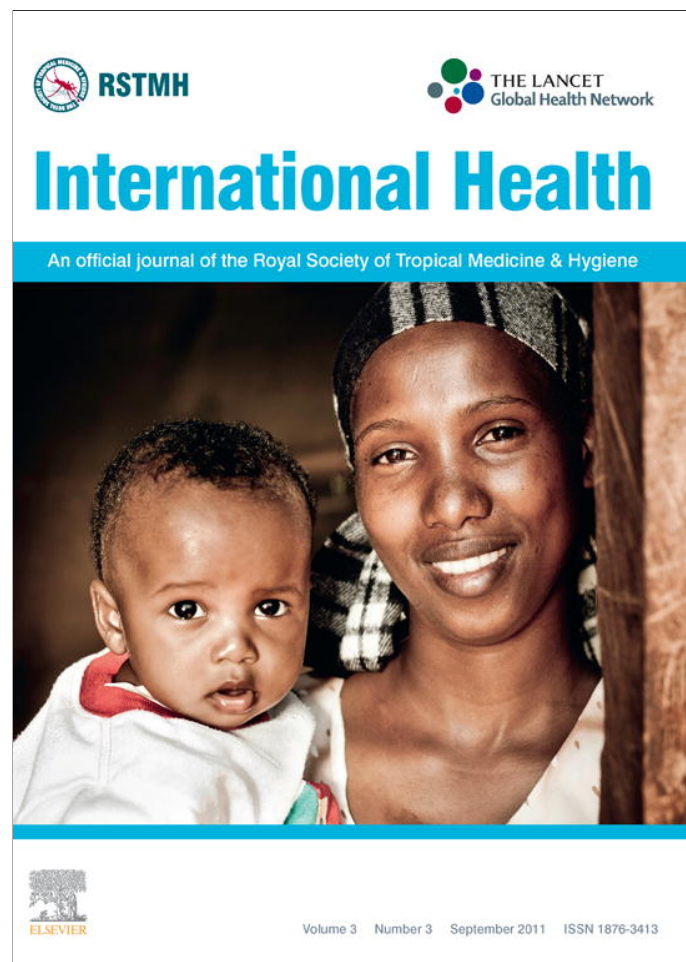


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

International Health

journal homepage: <http://www.elsevier.com/locate/inhe>

Review

Evaluating the scale-up for maternal and child survival: a common framework

Jennifer Bryce^{a,*}, Cesar G. Victora^{a,b}, Ties Boerma^c, David H. Peters^a, Robert E. Black^a^a Institute for International Programs, Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, 625 North Wolfe Street, Baltimore, USA 20205^b Post-Graduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil^c World Health Organization, Via Appia, Geneva, Switzerland

ARTICLE INFO

Article history:

Received 17 August 2010

Received in revised form 1 December 2010

Accepted 6 April 2011

Available online 14 May 2011

Keywords:

Evaluation Studies

Health Services

Health Status Indicators

Child Health Services

Child Survival

Maternal Health

ABSTRACT

Programs to reduce mortality among women and children are the target of new resources and redoubled commitment as the 2015 date for achieving the Millennium Development Goals approaches. The need for a common evaluation framework to guide the collection, analysis and synthesis of evidence is increasingly evident. This paper presents such a framework in four parts: (1) a conceptual model for the scale-up to MDGs 4 and 5 for maternal and child survival; (2) recommended indicators for each part of the model that bring together the work of various existing technical groups and prioritize a limited number of indicators for standardization and common use; (3) guidelines for documenting program implementation and contextual factors that may affect program implementation and its effectiveness in reducing maternal and child mortality; and (4) design considerations in evaluating the scale-up. We first present an overview of what is known and/or agreed upon within each of these areas, and in the discussion highlight areas of uncertainty or where there are gaps to be addressed.

© 2011 Royal Society of Tropical Medicine and Hygiene. Published by Elsevier Ltd. All rights reserved.

1. Introduction

There is broad consensus that a common evaluation framework focused on the health Millennium Development Goals (MDGs) is needed urgently.^{1,2} Countries will benefit because the core indicators and measurement approaches will be accepted by all donors and initiatives, reducing the need for separate measurement and reporting. The global community, including women and children, will benefit because the use of a common evaluation framework will generate more learning by supporting comparisons of results and costs within and across

countries that may use different approaches to implementation. The use of a common evaluation framework can stimulate recognition of methodological issues that need to be addressed and provide the foundation for harmonized research studies designed to provide state-of-the-art answers.

We propose such a framework here, building on and adapting the work of the International Health Partnership (IHP+),³ the WHO Toolkit on monitoring health systems strengthening⁴ and related work on developing strong country monitoring and evaluation platforms for national health strategies.⁵ Through these initiatives, countries and partners have agreed on a set of principles to guide evaluations (Box 1) and shared definitions of terminology to be used in discussions of evaluation efforts and findings. Our framework is specific to the scale-up of efforts to improve maternal, newborn and child health (MNCH) and achieve

* Corresponding author. Permanent address 2081 Danby Road, Ithaca, New York, 14850, USA. Tel.: +1 607 277 9731; +1 607 280 4800.

E-mail addresses: jbrycedanby@aol.com, jbryce@jhsph.edu (J. Bryce).

Box 1. Guiding Principles

- 1 Collective action: primary focus on the contribution of the collective efforts to scale-up the health sector response in countries.
- 2 Alignment with country processes: build upon national processes that countries have established to monitor and evaluate progress in the implementation of national plans.
- 3 Balance between country participation and independence: driven by country needs but conducted in a manner that maintains independence of evaluation.
- 4 Harmonised approaches: common protocols and standardized outcome indicators and measurement tools, with appropriate country adaptations.
- 5 Capacity building and health information system strengthening: systematic involvement of country institutions.
- 6 Adequate funding: between 5% and 10% of the overall scale-up funds set aside for monitoring performance, evaluation, operational research and strengthening health information systems.

MDGs 4 (child mortality) and 5 (maternal mortality and access to reproductive health).⁶ It was developed initially for use by donors and countries working with the Catalytic Initiative to Save One Million Lives, but has since been adopted and refined for use in the Countdown to 2015 for Maternal, Newborn and Child Survival⁷ and other initiatives.

The framework has four parts, each of which is addressed in subsequent sections: (1) a conceptual model for interventions and strategies aimed at improving maternal and child survival; (2) a set of core indicators to be used in monitoring and evaluation of progress; (3) guidelines for documenting program strategies and inputs; and (4) recommendations intended to lead to a set of compatible designs for the evaluation of country-level initiatives, to allow comparisons across places and time. In the discussion we describe gaps and limitations in the framework and identify priority actions needed to realize potential for global learning.

2. A conceptual model for the scale-up

The purpose of a conceptual model is to delineate the pathways through which programmatic efforts, or 'inputs', are implemented through processes to achieve intermediate outputs and outcomes that lead to reductions in maternal and child morbidity and mortality and improvements in nutritional status. These relationships are illustrated in [Figure 1](#). This simplified model builds upon those widely used in studies of maternal and child health,^{8–10} but reflects new evidence and emphasis in areas such as health systems strengthening and financial flows to MNCH programs that have not been incorporated previously. Equity considerations and contextual factors that may affect progress in these pathways are included at the bottom of the [Figure](#).

This model is generic, and must be modified to reflect the specifics of each program or initiative. First, the generic model must be adapted to reflect the contextual factors and MNCH program plans and expectations in specific settings – whether national or at state or district level. Second,

the model must be refined for specific applications such as to guide the design and analysis of research and evaluation studies. This generic model provides only a starting point for the evaluation research needed to understand better the relationships between specific types and levels of inputs and processes and results in terms of outputs, outcomes and impact. Evaluations may also need to account for more complex pathways to scaling up health services, which may involve feedback loops, emergent behavior, or phase transitions, which are typical of complex adaptive systems. These evaluations can be informed by conceptual frameworks that take these pathways into account, particularly to examine the diffusion of innovations or how health services organizations learn.^{11–13} Three specific applications include using the model as a guide for program planning, as a basis for designing prospective program evaluations, and as a road map for program-relevant research.

3. Standard indicators for evaluating the scale-up

The term indicator is used here to mean a quantitative measurement that produces results that are comparable across various contexts and over time. Our proposals for standard indicators are guided by a set of criteria tied to the overall goal of contributing to global learning. To the extent possible, each indicator should be valid (both accurate and reliable), meaningful in that it addresses an important part of the conceptual model describing pathways to impact, feasible for measurement and consistent with global standards where they exist.

All aspects of the conceptual model are not equally amenable to quantitative measurement using a set of standard indicators. MNCH program efforts are diverse, with varying types and levels of inputs, often using innovative processes and program strategies. The common evaluation framework therefore includes both indicators and qualitative and quantitative documentation to ensure the results produced under the common evaluation framework can be understood and interpreted.

An initial set of standard indicators is proposed here, along with discussions of data sources and methods. Areas in which further research and/or consensus building is needed are summarized in the discussion section.

4. Standard indicators for MNCH impact

For the purposes of the common evaluation framework, [Figure 1](#) proposes that impact be defined as changes in health status, and more specifically, as changes in morbidity, mortality and nutritional status. The measurement unit for impact is the individual; data would ideally be obtained through routine health information systems but *de facto* in most countries with high levels of maternal and child mortality are collected through interviews conducted in representative samples of households.

Five standard indicators of MNCH impact are proposed at present: under-five mortality, neonatal mortality, distribution of under-five deaths by cause, and the prevalence of stunting and wasting. The under-five mortality rate (the probability of dying before five years of age) is perhaps the

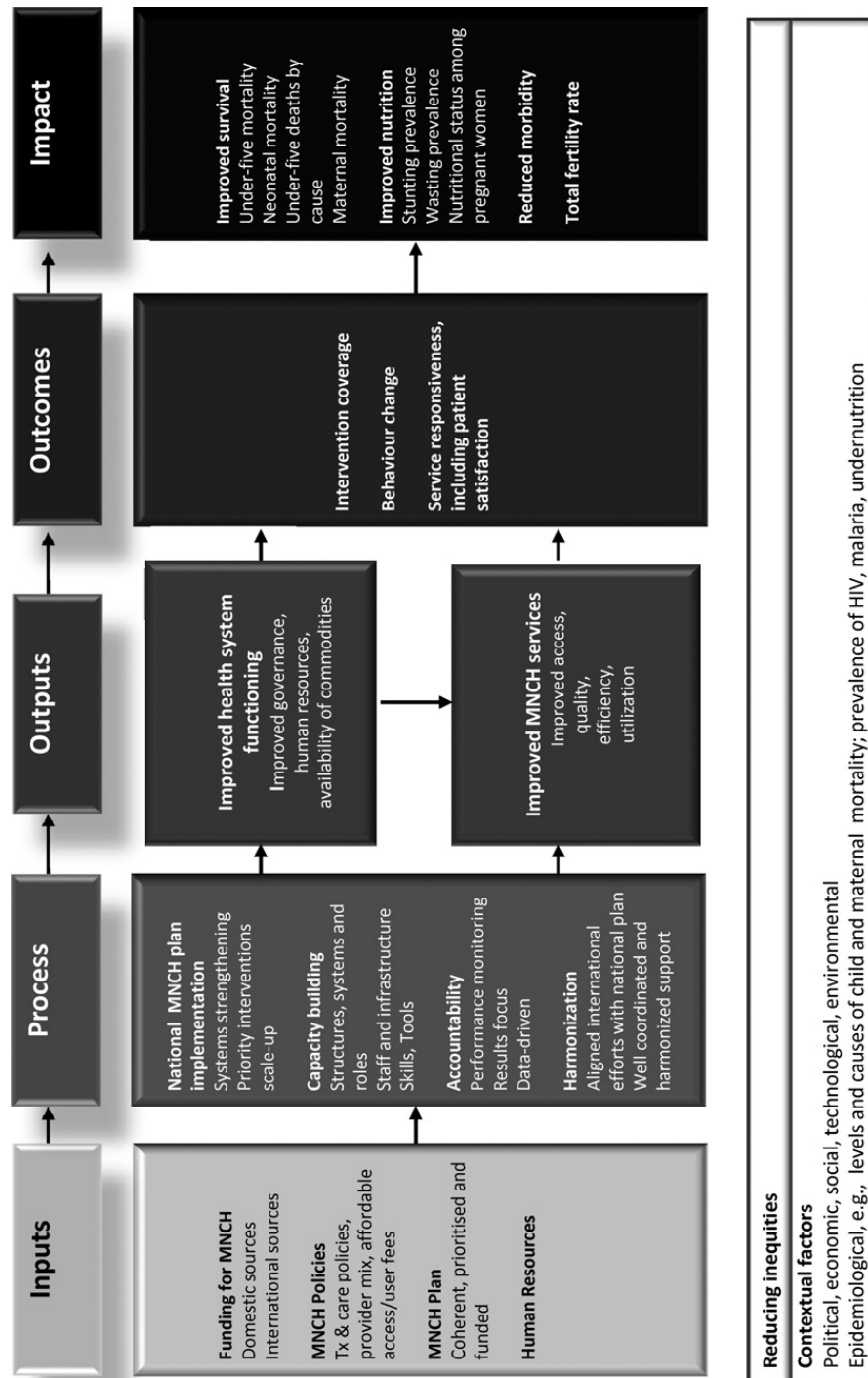


Figure 1. A conceptual model for the scale-up to MDGs 4 and 5.

single most widely-used measure of population health, and is one of the global indicators used for tracking progress toward the MDGs. Under-five mortality is considered the most robust measure of child mortality based on retrospective data collected in surveys and censuses.¹⁴ The neonatal mortality rate (deaths in the first four weeks of life per thousand live births) is progressively gaining importance as newborn deaths account for an increasing proportion of under-five deaths and are in general due to different patterns of causes. This indicator is recommended rather than the more traditional infant mortality rate (the number of deaths of infants under one year old in a given year per 1000 live births in the same year) because the latter is subject to heaping of deaths at 12 months when measured through direct birth histories. There is growing recognition that information on the cause of death is needed as a basis for assessing intervention effectiveness and informing programmatic decisions. Cause of death profiles for children under the age of five are changing rapidly, and will continue to do so, as a result of rapid development and associated secular trends and of widespread implementation of effective interventions to prevent specific diseases.¹⁵ Stunting is a contributing cause of about a quarter of child deaths, and severe acute malnutrition (wasting) has a high case fatality that can be addressed through timely treatment of infections and therapeutic feeding. Stunting and wasting should be included as standard indicators because they are important indicators of overall child health, and MDG 4 is unlikely to be achieved without associated improvements in child nutritional status. Nutritional status is also an important contextual factor; for example, famine is likely to undermine health gains achieved through other interventions. Standard consensus indicators for other impact elements are not yet available and are identified in the discussion as a priority area for further work.

5. Standard indicators for MNCH outcomes

The types of outcomes included in the framework (Figure 1) refer to measurements of effects in a beneficiary population, and include intervention coverage, behaviour change and increased service responsiveness to population needs. Recommended standard indicators for these areas are presented in Box 2.

5.1. Intervention coverage and behavior change

Coverage is defined as the proportion of individuals who need a service or intervention who actually receive it. Behavior change can be assumed to result from receipt of an intervention and standard indicators for these two areas are therefore considered together. The measurement unit for coverage and behavior change is the individual or its family; again data could and eventually should be collected via routine systems but *de facto* in low- and middle-income countries are obtained through interviews conducted in representative samples of households or (for a few interventions including prevention of mother to child transmission of HIV (PMTCT), vitamin A supplementation and vaccinations) through the combination

Box 2. Standard indicators for MNCH outcomes

Intervention coverage and behaviour change^a

Nutrition

1. Exclusive breastfeeding (<6 months)
2. Breastfeeding plus complementary food (6-9 months)

Child Health

3. Vitamin A supplementation coverage
4. Measles immunization coverage
5. DPT3 immunization coverage
6. HiB3 immunization coverage
7. Oral rehydration and continued feeding
8. Insecticide-treated net coverage
9. Antimalarial treatment
10. Prevention of mother-to-child transmission of HIV
11. Care seeking for pneumonia
12. Antibiotic treatment for pneumonia

Maternal and Newborn Health

13. Contraceptive prevalence
14. Unmet need for family planning
15. Antenatal care (at least one visit)
16. Antenatal care (four or more visits)
17. Neonatal tetanus protection
18. Intermittent preventive treatment of malaria in pregnancy
19. Skilled attendant at delivery
20. C-section rate
21. Timely initiation of breastfeeding
22. Postnatal care for mothers
23. Postnatal care for babies who were born at home

Water and Sanitation

24. Use of improved drinking water sources
25. Use of improved sanitation facilities

Inequities in Services

Breakdown of the above indicators by gender, urban/rural residence, wealth quintiles and regions of the country.

^a Taken from Countdown to 2015 for Maternal, Newborn and Child Survival, and to be updated as new evidence and data become available to remain consistent with future Countdown cycles (<http://www.countdown2015mnch.org/>).

of program reports and household survey data. We propose standard coverage indicators for those interventions that are supported by evidence demonstrating impact on maternal and child mortality or nutritional status in low- and middle-income countries. Countdown to 2015 (<http://www.countdown2015mnch.org/>) has identified a set of standard consensus coverage indicators accepted by United Nations Agencies and development partners, which we propose be adopted and measured in standard ways in all settings where the relevant interventions are being implemented.¹⁶

5.2. Responsiveness

WHO has defined service quality as one element of a broader outcome of health system 'responsiveness';⁴ other components of responsiveness included service prompt-

ness, access to social networks and communication.¹⁷ Measures of timeliness, social networks and quality of communication are important and are appropriately included in many monitoring and evaluation plans, but do not meet the criteria for standard indicators as defined here. Patient satisfaction is measured in many settings and appears likely to meet the criteria for relevance, susceptibility to change and feasibility, but indicators used to date have limited cross-cultural validity. Some information in these areas may be considered as part of documentation and are addressed in that context below; the need for better measures in this area is highlighted in the discussion section of this paper.

6. Standard indicators for MNCH outputs

The common evaluation framework includes two categories of MNCH outputs: improved MNCH services and improved health systems functioning, both of which measure factors on the supply side of the health system. At present there are no widely-used standard indicators in these areas, although considerable work is under way and the WHO Toolkit⁴ provides a useful starting point. Priorities for research in this area are proposed in the discussion.

7. Standard indicators for inputs and processes

Developing quantitative measures of MNCH program processes (Figure 1) is difficult because of the diversity of approaches and methods used across programs and settings. Much of this information is more suited to documentation, described in the next section.

For program inputs, however, there are standard indicators that should be considered as part of the common evaluation framework. At a minimum, efforts should be continued and expanded to track resource flows to national and sub-national levels from all sources (overseas development assistance, public and private), expressed as the amount disbursed *per capita* or per target population.¹⁸ The total and incremental costs of each program strategy *per capita* should also be measured, taking into account the fact that the scale-up of MNCH programs may change the way that people behave. Measures of the adequacy of human resources, such as the number of doctors, nurses and midwives per 10 000 people,¹⁹ are unlikely to be useful across contexts with varying population and health system characteristics.

8. Reducing inequities

There is growing concern that programs and interventions fail to reach those who need them most, and that overall progress in health indicators can hide important gender, socioeconomic or ethnic group differences. Assessing the impact of a program on equity is thus becoming more and more important. The conceptual model (Figure 1) reflects the importance of addressing equity across the entire pathway from inputs to impact, and the need to stratify outcomes (coverage) and impact results by population subgroups. Recent progress in measurement of socioeco-

omic position and in equity measurement tools allows the routine incorporation of this dimension in surveys and other data sources.²⁰

9. Documenting program implementation and contextual factors

Documentation has traditionally been overlooked as a key component of program evaluations. However, the quantitative data produced by programs and evaluation teams for specific components of the common evaluation framework will have limited usefulness unless complementary information is available about how specific programs were implemented. Such information is particularly useful for understanding why a given program had an impact, or failed to do so. It is also crucial to document key characteristics of the local context such as demographic patterns, socio-economic factors and the presence of other child survival and global health initiatives that can be used to identify potential confounders and effect modifiers for analysis and interpretation.²¹ Much of the information needed for the documentation of contextual factors is likely to be available at country level; programs can draw on this information rather than developing parallel systems. Box 3 presents some preliminary proposals on documentation. A standard protocol has been developed for programs supported by the Catalytic Initiative; these may provide a useful starting point for others.²²

10. Evaluation design issues

The overall aim of evaluations of public health programs is to judge their effectiveness, costs, and other consequences in order to guide decisions about continuing, expanding, or changing programs, either on their own or in the context of other uses of public resources available to improve public health. To accomplish this aim, evaluation should be integrated in all stages of a program starting with the design including baseline assessments and determination of historical trends, and ending with the assessment of ultimate outcomes and impact.²³ Within this overall aim, however, there are a range of purposes for evaluation that are often less explicit, but hold important implications for design.²⁴

The common evaluation framework should be applied using the strongest designs available – whether for new data collection or for secondary analysis of existing data sets. For prospective evaluations, randomization will rarely if ever be possible (or even desirable, given that the goal of the scale-up is universal coverage with interventions already proven effective). As support for the scale-up expands, ‘virgin’ comparison groups without programs intended to reduce mortality among women and children are increasingly rare. Under these circumstances all evaluations must document carefully all inputs to the program, by source, and keep track of contextual factors that are likely to affect impact in both the intervention areas and, where they exist, in the comparison areas.

Because programs are almost never deployed with equal intensity in different regions of a country, designs that track program implementation in a stepwise man-

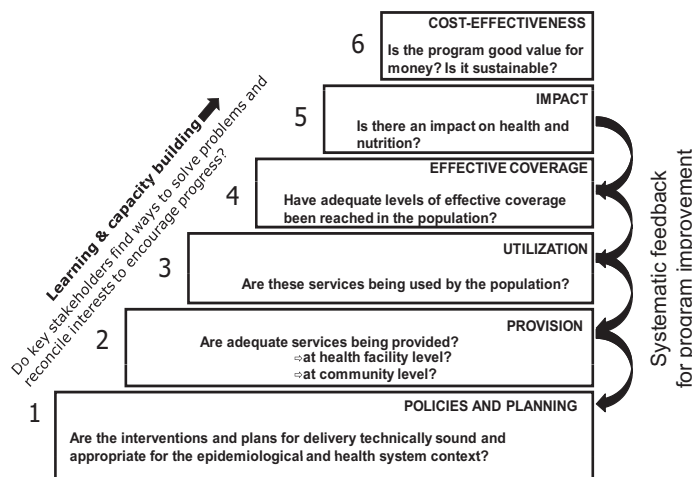


Figure 2. A stepwise approach to evaluating the scale-up to MDGs 4 and 5.

Box 3. Documenting program implementation and contextual factors

Basic information on the program

Interventions to be scaled-up, and how they are delivered (e.g., through facilities, at community-level)
 Target population; geographic scope
 Human resource inputs: numbers trained and additional staff recruited standardized to population of 10 000
 Incentives related to the program
 Financial inputs
 Infrastructure and equipment

Key contextual factors

Health Sector

Demographic (population size and density, urbanization, fertility rate, family size, ethnic and language groupings)
 Epidemiologic (epidemics and outbreaks, health transition)
 Health systems (health facility coverage, size and role of private providers, management information systems, referral systems, user fees, health insurance schemes if in place, health policies)
 Other health programs, activities likely to have an impact on population health.

Other

Political (change in regime, role of state in health sector and economy, level of government decentralization, processes for public involvement in decision-making, relevant labor laws, tax policies)
 Economic (economic growth, inflation, exchange rates, poverty headcount)
 Social (cultural aspects affecting demand for services, women's education and literacy, civil conflict)
 Technological (technological change [e.g. mobile phones, internet, diagnostics], factors affecting barrier to entry or diffusion of technological innovations)
 Legal (laws to protect consumers, discrimination law, laws to enforce contracts and regulations, rule of law)
 Environmental (climate and climate change, natural disasters, famine)

ner (see Figure 2) are strongly recommended, even in the absence of randomization. In this evaluation approach, intermediate findings on the pre-requisites for impact (e.g., sound program policies and program design, adequate provision and utilization of services, minimum service quality and population-based coverage) can be used to improve program effectiveness and head off costly impact evaluations in settings where they are not warranted due to poor implementation. A fuller discussion of these issues, the potential of time-series designs and econometric approaches, and a proposal for developing national evaluation platforms that link existing and new data sources at district level for use in evaluating alternative approaches are available elsewhere, including examples of possible designs.²⁵

11. Discussion

In this paper we have tried to bring a mosaic of guidelines and indicators together into a practical framework for evaluating the scale-up to MDGs 4 and 5, providing clear recommendations and establishing a foundation for global learning. There are gaps – areas where what to measure or how to measure are not clear or there is no consensus, and other areas that require additional research or conceptualization. In this section we highlight specific areas where further work is needed, and suggest immediate next steps.

The first gap that hinders full implementation of a common evaluation framework is the need for consistent, comparable data on framework elements that are available in most if not all countries but particularly those with high levels of maternal, newborn and child mortality. Some of the most important areas of work needed to address this gap are highlighted below.

First, there is a need to develop consensus on measures of population health impact for reproductive and maternal health programs. Country-specific maternal mortality ratios are used by some initiatives, e.g., Country Health Systems Surveillance (CHeSS)⁵ but the need for large sample sizes makes it difficult to assess short-term trends and all estimates have large uncertainty margins. The adolescent

birth rate, coverage of skilled attendance at birth (or institutional deliveries), unmet need for family planning and the contraceptive prevalence rate are potential proxies for impact in these areas.

Second, there is a need to improve existing indicators of coverage, particularly those for the treatment of childhood illnesses in which the validity of neither the denominator (e.g., cases of childhood pneumonia) nor the numerator (proportion of children with pneumonia who receive effective antibiotic treatment) have been validated. The Child Health Epidemiology Reference Group is working to address these issues with support from the Bill & Melinda Gates Foundation,²⁶ and the results of this work will be reviewed by Countdown to 2015 (<http://www.countdown2015mnch.org/>) and incorporated into the common evaluation framework as they become available.

Third, there is a need to test additional outcome indicators to provide valid and timely results that can be compared across settings and over time in the areas of health system responsiveness, the technical quality of service delivery compared to standards and coverage for individual interventions subsumed within broader indicators of service contact such as antenatal or postnatal care.

Fourth, there is a need to forge a wider consensus on a small number of indicators of health systems access, and capacity, and service quality that take into account the constraints of weaker health systems with poor data. The CHES activity sponsored by IHP+ is working to forge consensus on a reduced set of standard indicators for health systems strength; the framework will need to be updated as this work continues and the WHO Toolkit may serve as a useful central repository for this learning.

Greater focus is also needed on documenting inputs and processes, and in refining broad concepts such as 'capacity building' to focus on a hierarchy of levels reflecting how various forms of capacity building link to one another in an interactive process.²⁷ Measuring costs is a particular challenge given the need to disaggregate joint costs to reflect inputs to specific service programs and because most countries lack capacity to conduct costing studies on a regular basis. Only through joint efforts to define and track inputs, processes and contextual factors will we be able to learn about the types of strategies that work in specific contexts and their human and financial costs.

A second important gap is in finding the appropriate balance between efforts to strengthen health systems overall while maintaining an appropriate focus on the specific improvements required to meet MNCH program needs. Health systems by definition are characterized by joint production functions, and the same inputs (e.g., infrastructure, health workers) are needed to produce many different outputs and outcomes. Selective strengthening of health services for specific MNCH outcomes is therefore in some areas illogical, and yet not all health policies and health systems strengthening activities are equally important or feasible in a given context. Ongoing discussions around the 'six building blocks' framework developed by WHO, e.g., Atun, R. et al.¹¹ and the development and testing of indicators of health systems strength must continue with the urgent aim of defining a com-

mon framework for health systems strengthening that articulates with this framework and the broader IHP+ framework from which it was adapted. At the same time, MNCH programs and their evaluation counterparts can begin with the proposals in the WHO Toolkit and work to assess their utility in field settings with subsequent refinements.

Finally, this framework is based on a public sector perspective, and the private sector plays a large and growing role in health services in many (although not all) countries with high burdens of maternal, newborn and child mortality. Further work is needed to address the private sector in ways that will promote monitoring and evaluation and ensure it is appropriately represented in the global learning agenda.

The common evaluation framework can contribute to global learning and improved programs for women and children, but only if it is widely used and continues to be refined through experience and research. What is clear now is that all evaluations must be technically sound and independent of program implementation or advocacy. They should involve in-country research institutions, with support – if necessary – from external institutions with demonstrated expertise in impact evaluation. There should be a systematic plan for review and discussion of evaluation results as they become available to provide opportunities for regular and open interaction between evaluators and program implementers. This would occur from the start of the evaluation, when the conceptual model is established and agreed to, through evaluation design, feedback of intermediate and final results, and the interpretation of results and determination of their implications for evaluation. The joint assessments and harmonization activities being used by IHP+ hold promise in this area, and may help to strike the right balance between a systems approach to monitoring and evaluation and the need to maintain focus on MNCH-specific issues. Finally, there must be resources committed to evaluation.

The time has come for all governments and development partners to adopt, use, and together refine a common evaluation framework for the scale-up to the MDGs. This work has already begun; here we present an adaptation of the framework specific to the MDGs related to maternal, newborn and child survival. Further work is needed urgently to develop, test and build consensus around additional elements of the framework, including indicators for critical parts of the pathways related to reproductive and maternal health, health system responsiveness and the documentation of inputs and processes. Existing indicators must be the starting point for intensified efforts to improve measurement. Finding the appropriate balance between overall health system strengthening and a focus on those parts of the system that are most important for MNCH programs remains a challenge. Independent evaluations that are based on a common framework and that are technically sound will benefit women and children by contributing to improved programming and the evidence needed to generate additional resources and expand their impact.

Authors' contributions: All authors contributed to conceptualizing the framework. JB and RB wrote the first draft;

JB wrote subsequent drafts with substantive contributions from all authors. All authors approved the final version for submission. Jennifer Bryce will serve as guarantor for this paper.

Acknowledgements: This paper reflects contributions from many individuals working with the Institute for International Programs at the Johns Hopkins Bloomberg School of Public Health. The working paper prepared in June 2008 included sections developed by David Peters and Joe Naimoli of the World Bank on health systems and Damian Walker, Marjorie Opuni-Akuamo and Ben Johns on costs, and numerous reviews and discussions with those involved in evaluations of large-scale public health programs (Agbessi Amouzou, Abdullah Baqui, Kate Gilroy, Linda Bartlett, Gareth Jones). The work on documenting program implementation and contextual factors is led by Kate Gilroy and Elizabeth Hazel. Proposals for standard indicators reflect work developed for the Doris Duke Charitable Foundation's African Health Initiative, with contributions by Jennifer Harris Requejo, Lawrence Moulton and Ben Johns.

Funding: Funding was provided by Canadian CIDA and the Bill & Melinda Gates Foundation through their support to IIP-JHU for the independent evaluation of the Catalytic Initiative to Save One Million Lives, and by the Doris Duke Charitable Foundation through their support to IIP-JHU as the Coordinator for the African Health Initiative Data Collaborative.

Conflicts of interest: None

Ethical approval: Not required

References

1. Victora CG, Black RE, Bryce J. Learning from new initiatives in maternal and child health. *Lancet* 2007;**370**:1113–4.
2. Murray CJ, Frenk J, Evans T. The Global Campaign for the Health MDGs: challenges, opportunities, and the imperative of shared learning. *Lancet* 2007;**370**:1018–20.
3. Boerma T, AbouZahr C, Bos E, Hansen P, Addai E, Low-Beer D. Monitoring and evaluation of health systems strengthening: an operational framework. <http://www.who.int/healthinfo/HSS.MandE.framework.Oct.2010.pdf> [accessed 29 November 2010].
4. World Health Organization. Measuring health systems strengthening and trends: a toolkit for countries. WHO, November 2009. www.who.int/healthinfo/HSS.MandE.framework.Nov.2009.pdf [accessed 30 March 2011].
5. Country Health Systems Surveillance. Work Plan. [http://www.internationalhealthpartnership.net/CMS.files/documents/country_health_systems_surveillance.\(chess\).EN.pdf](http://www.internationalhealthpartnership.net/CMS.files/documents/country_health_systems_surveillance.(chess).EN.pdf) [accessed 29 November 2010].
6. United Nations. The Millennium Development Goals report 2009. New York, NY, USA: United Nations Department of Economic and Social Affairs, 2009.
7. Bhutta ZA, Chopra M, Axelson H, Berman P, Boerma T, Bryce J, et al. Countdown to 2015 decade report (2000–10): taking stock of maternal, newborn and child survival. *Lancet* 2010;**374**:2032–44.
8. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al., for the Maternal, Child Undernutrition Study Group. Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet* 2008;**371**:243–60.
9. Mosley WH, Chen LC. An analytical framework for the study of child survival in developing countries. *Popul Dev Rev* 1984;**10**(Supplement):25–45.
10. Dahlgren G, Whitehead M. *Policies and strategies to promote social equity in health*. Stockholm: Institute of Futures Studies; 1991.
11. Atun R, de Jongh T, Secci F, Ohiri K, Adeyi O. Integration of targeted health interventions into health systems: a conceptual framework for analysis. *Health Policy Plan* 2010;**25**:1–14.
12. de Savigny D, Adam T, editors. *Systems thinking for health systems strengthening*. Geneva: Alliance for Health Policy and System Research and World Health Organization; 2009.
13. World Bank. What is a health system? In: *Healthy development: The World Bank strategy for health, nutrition and population results*. Washington, DC: The World Bank, 2007, p. 173–8.
14. Hill K. Approaches to the measurement of child mortality: a comparative review. *Popul Index* 1991;**57**:368–81.
15. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* 2010;**375**:1969–87.
16. Countdown to 2015 for Maternal, Newborn and Child Survival. Countdown to 2015 Decade Report (2000–2010). <http://www.countdown2015mnch.org/documents/2010report/Countdown-ReportAndProfiles.pdf> [accessed 30 March 2011].
17. WHO. Meeting on Responsiveness Concepts and Measurement. Geneva, Switzerland, 13–14 September 2001. Publication No. HFS/FAR/RES/00.1. http://www.who.int/health-systems-performance/technical_consultations/responsiveness_report.pdf [accessed 28 June 2010].
18. Greco G, Powell-Jackson T, Borghi J, Mills A. Countdown to 2015: assessment of donor assistance to maternal, newborn and child health between 2003 and 2006. *Lancet* 2008;**371**:1268–75.
19. World Health Organization. *World Health Report 2006: Working for health*. Geneva: World Health Organization; 2006.
20. O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. *Analyzing health equity using household survey data: a guide to techniques and their implementation*. Washington, DC: The World Bank; 2007.
21. Victora CG, Armstrong Schellenberg J, Huicho L, Amaral J, Arifeen SE, Pariyo G, et al. Context matters: Interpreting impact findings in child survival research. *Health Policy Plan* 2005;**20**:S1:i18–i31.
22. Gilroy K, Hazel E, and the Working Group on Evaluating the Catalytic Initiative. Guidelines for documenting program implementation and contextual factors in independent evaluations of the Catalytic Initiative. Working paper, V1.1. http://www.jhsph.edu/dept/ih/IIP/projects/catalytic_initiative/Documentation_guidelines_v1.pdf [accessed 28 June 2010].
23. Wynn BO, Dutta A, Nelson M. *Challenges in program evaluation of health interventions in developing countries*. Santa Monica, CA: Rand Corporation; 2005.
24. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 1999;**28**:10–8.
25. Victora CG, Black RE, Boerma JT, Bryce J. Measuring impact in the MDG era and beyond: A new approach to large-scale effectiveness evaluations. *Lancet* 2011;**377**:85–95.
26. Child Health Epidemiology Reference Group (CHERG). [place unknown]: CHERG; 2010. <http://cherg.org/home.html> [accessed 20 June 2010].
27. Potter C, Brough R. Systemic capacity building: a hierarchy of needs. *Health Policy Plan* 2004 Sep;**19**:336–45.