OXFORD

Lessons learned in using realist evaluation to assess maternal and newborn health programming in rural Bangladesh

Alayne Adams,[1,]* Saroj Sedalia,[2] Shanon McNab,[2] and Malabika Sarker[3]

[1]Centre for Equity and Health Systems, icddr,b, Dhaka, Bangladesh, [2]The Averting Maternal Death and Disability Program (AMDD), Mailman School of Public Health, Columbia University, New York, NY, USA and [3]James P. Grant School of Public Health, BRAC University, Dhaka, Bangladesh

*Corresponding author. Centre for Equity and Health Systems, icddr,b, GPO Box 128. Dhaka, 1000, Bangladesh.
E-mail: aadams@icddrb.com

Abstract

Realist evaluation furnishes valuable insight to public health practitioners and policy makers about how and why interventions work or don't work. Moving beyond binary measures of success or failure, it provides a systematic approach to understanding what goes on in the 'Black Box' and how implementation decisions in real life contexts can affect intervention effectiveness. This paper reflects on an experience in applying the tenets of realist evaluation to identify optimal implementation strategies for scale-up of Maternal and Newborn Health (MNH) programmes in rural Bangladesh. Supported by UNICEF, the three MNH programmes under consideration employed different implementation models to deliver similar services and meet similar MNH goals. Programme targets included adoption of recommended antenatal, post-natal and essential newborn care practices; health systems strengthening through improved referral, accountability and administrative systems, and increased community knowledge. Drawing on focused examples from this research, seven steps for operationalizing the realist evaluation approach are offered, while emphasizing the need to iterate and innovate in terms of methods and analysis strategies. The paper concludes by reflecting on lessons learned in applying realist evaluation, and the unique insights it yields regarding implementation strategies for successful MNH programming.

Key words: Bangladesh, implementation research, maternal and newborn health, realist evaluation

---

Key Messages

- Realist evaluation is a viable approach for evaluating the implementation of MNH programmes.
- The realist approach is appropriate for in-depth enquiry into theories driving programmes, programme fidelity, and explaining why programmes fail or succeed in different settings.
- A variety of qualitative methods provide information that can be used in realist evaluation.
- The strength of findings in realist evaluation depends greatly on high quality data, and framing findings in a manner that complements more traditional evaluation methods.

Introduction

In public health, evidence-based interventions often fail due to challenges in implementation (Sanders 2006; Porter and O'Halloran 2012). Although experts in the field of Maternal and Newborn Health (MNH) may agree on life saving technologies and clinical approaches, there is less consensus on how to deliver these inputs in a manner that achieves effective coverage. The complexities of transforming inputs into outcomes constitute a 'black box of implementation' that frustrates efforts to agree on best practices for putting interventions into action (Harachi *et al.* 1999; Sullivan *et al.* 2008; Astbury and Leeuw 2010). In this context, the need for research that focuses on implementation processes is increasingly recognized (Pawson and Tilley 1997; Kazi 2003; Brownson *et al.* 2009). The growing field of implementation research investigates the programme execution with the understanding that successful transferability of programme models requires insight on how an intervention actually works and the theories that drive it (Chen and Rossi 1983; Bourguignon *et al.* 2007). Unlike traditional impact evaluation approaches that establish whether change in outcomes can be directly attributed to an intervention (Astbury and Leeuw 2010), realist evaluation focuses on the processes and contexts of implementation that yield impact (Pawson and Manzano-Santaella 2012). By examining 'what works, for whom and why' (Pawson and Tilley 1997), insights are gained about the interactions between interventions, communities, implementers and health systems that make programmes more or less successful.

The principles of realist evaluation

Pawson and Tilley (1997) describe realist evaluation as a theory-driven approach that views programmes as 'theories in action'. The role of the researcher is to propose and refine hypotheses for generative causality (Ekstram 1992) that enable systematic thinking about what, why and how programme inputs bring about change in key outcomes, and how context shapes these processes. Initial hypotheses, framed as programme theories, are tested using context, mechanism, outcome (CMO) configurations, which describe how an intervention is expected to work for which (group of) actors and how. Contexts represent conditions needed for an intervention to trigger (or not) mechanisms, the causal processes that produce particular outcomes or effects, whether positive or negative (Pawson and Tilley 1997). Articulated together, they become a CMO configuration, which begins to describe which contextual elements and what mechanisms led to different outcomes. As new insights emerge from data collection and analysis, hypothesized relationships between CMOs are iteratively altered to reflect realities on the ground. Rearticulated 'CMO configurations' better capture the key factors and processes responsible for positive or negative programme outcomes based on evidence. In addition to programme evaluation, realist principles have also been applied to secondary data synthesis and review (Pawson *et al.* 2005; Dieleman *et al.* 2009; Greenhalgh *et al.* 2009; Marchal *et al.* 2010) and, more controversially, incorporated into the design of randomized community intervention trials (Hawe *et al.* 2004).

Designing and implementing realist evaluation

This reflection on the realist evaluation process emerged from an assessment of three MNH programmes that sought to identify 'optimal' implementation strategies for scale-up in rural Bangladesh (Box 1). A realist evaluation approach was chosen because of its sensitivity to context and its flexibility in terms of methods. The realist approach also allowed for a simultaneous exploration of programme fidelity—or the extent to which programmes are designed and implemented in a manner congruent with theory. The research was conducted from January to June 2012, with approvals from the respective Institutional Review Boards of Columbia University, USA and BRAC University, Bangladesh.

Despite the burgeoning literature on realist evaluation, there is limited guidance for researchers who wish to employ the approach (Van Belle *et al.* 2010). Realist evaluation does not prescribe a particular method of data collection, hence most literature focuses on principles, not protocols (Rycroft-Malone *et al.* 2010). Nevertheless, given growing interest in the approach, reflections on how it works in practice are useful. Drawing on lessons learned in undertaking a realist evaluation of MNH programming in Bangladesh, seven steps (Van Belle *et al.* 2010; Marchal *et al.* 2012) are proposed to help operationalize the approach, recognizing the need for iteration and creativity as suits the setting, programme and research team involved (see Figure 1).

Working collaboratively with UNICEF, the evaluation team identified six priority domains that cross-cut the three MNH programmes under consideration: local level planning, supply side gap

---

Box 1. Description of UNICEF's rural MNH programmes

This evaluation analyzes three distinct MNH programmes supported by UNICEF, but implemented by a variety of partners.

1. The Joint Government of Bangladesh-UN Maternal Newborn Health Initiative (MNHI): established in 2007, this programme is jointly run by UNFPA, UNICEF and WHO.
2. Improving Maternal, Neonatal and Child Survival (IMNCS): established in 2008, UNICEF has been supporting supply side improvements in MOHFW facilities, and linkages with BRAC, a large NGO involved in community-based health delivery and demand generation.
3. Maternal, Neonatal and Child Survival Interventions (MNCS): established in 2008, a partnership between UNICEF and implementing NGOs that targets hard-to-reach areas and indigenous populations.

All three programmes share the goal of improving maternal and newborn health outcomes by supporting efforts to promote healthy behaviours such as increased ante-natal care, timely referral, safe delivery and post-natal follow-up. Programmatic efforts to strengthen health systems were also undertaken, however, initiatives around local level planning, improved referral and social accountability vary between programmes.

Each programme employed different programme models and experimented with different tools. For example, some programmes promoted accountability by means of maternal death reviews while other programmes did so through community awareness raising. Different management approaches were also used across programmes, such as supervision, remuneration and programme-specific training and refresher training requirements for CHWs. UNICEF's role in each programme varied.
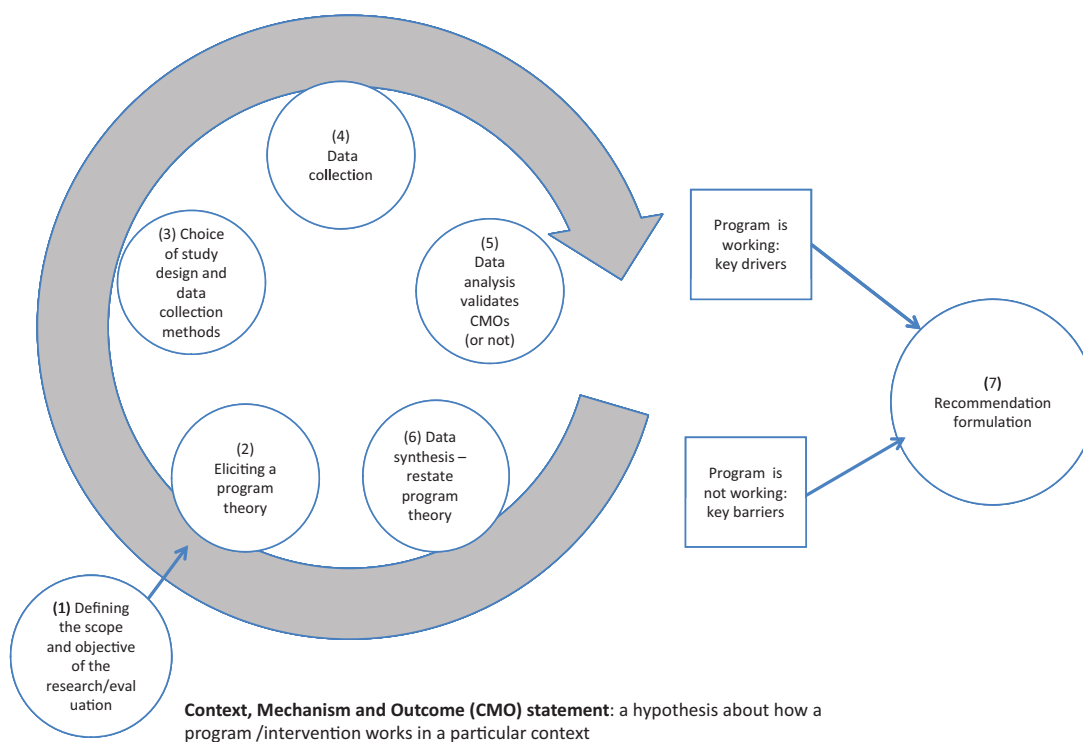
Context, Mechanism and Outcome (CMO) statement: a hypothesis about how a program /intervention works in a particular context

Figure 1. CMO statement: a hypothesis about how an programme/intervention works in a particular context

filling, human resources, Community Health Workers (CHWs), referral and accountability. For the purpose of this article, the example of CHWs is used to present and discuss the steps followed in this application of realist evaluation. All three MNH programmes deployed CHWs at the local level to visit pregnant and post-natal women and newborns in their homes, provide information on best maternal care practices and danger signs, and facilitate emergency transport to a health facility, with the understanding that women would listen to and implement the advice provided. As later discussed (Step 2), while the implicit programme theory underlying each of the programme models was similar, they were implemented in different ways, with variation around the frequency and scope of training, opportunities for supportive supervision, incentive systems and levels of support for hospital transfer.

Lessons learned in using realist evaluation

(1) Defining the scope and objectives of the research/evaluation

Prior to initiating a realist evaluation, it is worthwhile to consider the needs of those commissioning the research. Realist evaluation offers a unique perspective on programmes and how they work, grounded in theory. It provides in-depth insight that is complementary to more traditional impact or process evaluations, but is time and resource intensive. UNICEF expressed interest in understanding the attributes of successful MNH implementation, with the aim of identifying an optimal model for MNH programming. The organization's questions seemed appropriate to realist evaluation as they wished to understand what made interventions work in different settings. After discussion, a consensus was reached to explore optimal strategies for implementation, with less emphasis on identifying an 'optimal' programme model. There was, however, an expectation that the evaluation produce recommendations for action that, on reflection, were at odds with a theory-focused approach.

Once the scope and objectives of the research were decided, a research team was formed with wide expertise in content and methods, including anthropologists, public health researchers, general clinicians and MNH experts. This diverse research group ensured that the focus on MNH implementation went beyond the assessment of clinical outcomes, and took into account the processes and systems that enable outcomes to be achieved.

(2) Eliciting a programme theory

As preparation for programme theory articulation, the research team collected and reviewed all existing programme documentation including impact evaluations, mid-term reports and monitoring data; and conducted field visits and key informant interviews (KIIs) with donors and implementing agencies. This familiarization process oriented the research team to the programme models UNICEF had chosen to achieve intended outcomes, the contexts in which they were being implemented, and the approaches that implementing partners used to put programmes into action.

Realist evaluation views programmes as theories put into practice, and is intended to test and refine programme theories while explaining how the theories operate in real world settings. Articulating programme theory is thus the starting point of any realist evaluation. Programme theory is 'a set of statements that describe a particular programme, explain why, how and under what conditions programme effects occur, predict the outcomes of the programme, and specify the requirements necessary to bring about desired programme effects' (Sidani and Sechrest 1999). In reality, however, these conjectures are rarely explicitly articulated by programmes and implementing organizations. Indeed, none of the three MNH programmes under investigation had clearly stated programme theories. Step 2 of the realist evaluation process, therefore, involved surfacing often implicit assumptions underlying

the intervention design, with the goal of articulating programme theory.

A strategy helpful to this process was a 'problem/solution' thought exercise (Box 2) that was undertaken within the research team, and later with UNICEF. Staff turnover typical of international organizations like UNICEF, sometimes impeded productive sharing as institutional memory was limited. Many original UNICEF staff members had since departed, and limited documentation was found that explicitly described programme theory. Subsequent efforts to construct programme theory, therefore, represent informed conjectures of what programme implementers were trying to accomplish.

Preliminary fieldwork was performed in several districts to provide greater contextual familiarity among team members, and prepare for CMO articulation. Anthropologists shadowed CHWs and health facility staff, observing the working environment. The research team, which included five masters level anthropologists, used guidelines to elicit narratives about women's experience with the health system during pregnancy, and performed structured facility observations to capture dynamics of care, patient flow and setting over a period of 2 weeks. These contextual data, as well as findings from document reviews and site visits informed the development of CMOs. Table 1 presents a series of critical questions for CMO articulation, each of which is illustrated using the example of CHWs. Working through this series of questions, the underlying contexts and mechanisms at work within each programme domain were articulated. The specification of mechanisms as distinct from activities was particularly challenging, yet critical (Astbury and Leeuw 2010). Drawing on the example of CHWs, while training on key MNH messages was an important programme activity, the mechanism affecting CHW performance was the acquisition of knowledge of key MNH messages. If knowledge is not imparted, the intervention fails. Although the questions in Table 1 were helpful in articulating CMOs, there was still concern whether mechanisms were being adequately captured and conceptualized. For example, what aspect of the interaction between implementers and CHWs motivates CHWs to take their role seriously? The distinction between processes/activities and mechanisms was often ambiguous, resulting

in some CMOs that fell short in capturing the subtleties of programme theory.

(3). Choice of study design and data collection methods

Study design
Omissions in the formulation of CMOs relating to incomplete understanding of context, poorly articulated mechanisms and unknown outcomes, were recognized by the research team, and precipitated a return to existing project documentation to clarify elements of programme theory. These elements, particularly mechanisms (Table 2), then became a basis for identifying the data that would be needed to test CMOs, and the range of methods and tools best suited for data collection. Importantly, an appreciation of the complexities of capturing mechanisms of change emphasized the need to develop tools that explored more qualitative dimensions of behaviour such as CHW motivation and comfort navigating the health system. Research tools deepened understanding of context and processes around and within programmes (e.g. what do CHWs like about their training?) and generated new knowledge in areas where there was none (e.g. what do CHWs want from supervisors?).

A case study research design was adopted due to its strengths in exploring and generating theory, its methodological flexibility, and its ability to provide insight through detailed contextual analysis of similarities and differences (Yin 2003). Each MNH programme became a case, and study methods were designed to explore how CHW interventions were working across cases, and how this corresponded to shared programme theory that CHWs link women to services through home visits. Study methods and tools would therefore need to elicit understanding of how and why CHWs in the different MNH programmes were motivated to perform this particular function and to gauge women's experiences with CHWs and the maternal health services they provided.

At this stage, the research team recognized that the breadth of inquiry was too wide to fit within agreed timelines. Because UNICEF requested actionable recommendations for programme implementation, the research team decided to focus more on those working within

---

Box 2. Developing programme theory for CHWs
The problem
Despite considerable progress in MNH intervention coverage over the last decade, the use of formal MNH services in rural Bangladesh remains sub-optimal. According to data from the 2011 Bangladesh Demographic and Health Survey, only one-quarter of women receive four or more antenatal care visits, and a medically trained provider attends one-third of births. Contextual factors such as weak referral networks, lack of trust in the formal health care system, financial consequences of care seeking, and widespread and persistent misconceptions regarding maternal and newborn care practices hinder optimal MNH. These contextual and financial barriers are even more pronounced among the ultra-poor, and those living in hard-to-reach areas where distance to services and cultural barriers are pronounced. In general, functional systems of routine identification of pregnancy and danger signs, and referral are not in place. Government community-health workers charged with these tasks may lack the necessary skill set and supervision to provide them and thus have relied on NGOs and UN agencies for support to provide these services.
The programme theory to address the problem
Bangladesh has had a long and successful experience with local field-level health workers disseminating key health messages (e.g. family planning, oral rehydration therapy). In keeping with this tradition, UNICEF is working with partner NGOs to support the development of a cadre of locally recruited CHWs whose role it is to raise MNH awareness by means of house-to-house visits and community mobilization.
CHWs promote increased MNH coverage by visiting women in their homes, and delivering health messages around MNH. These messages help empower women to seek health services at the time of delivery. CHWs are motivated to fill this role by means of training and supportive supervision. Further, women's reluctance and distrust of facilities will dissipate because of trust in her CHW, and the social support she provides.

Table 1. Key questions to elicit CMOs for the CHW domain

Question	Example from study
1. What is the problem for which this programme is the solution and what will be different as a result? (outcomes)	Women do not feel/know that they have options for maternal healthcare and distrust facilities.
2. How does this programme solve the problem? (mechanisms of change)	Investment in the capacity of CHWs to link women to the care that they need.
3. What are the specific programme strategies and organizational structures used to implement this solution? (planned vs. actual programme implementation)	Planned implementation involves deployment and support of CHWs, however, actual implementation reveals inadequate supervision and monitoring.
4. What about the strategies or structures makes people change their behaviour? i.e. how do CHWs influence women's behaviour? (mechanisms)	CHWs are trained to deliver MNH messages and offer support during regular household visits. This puts mechanisms into action such as generating knowledge among women, or creating trust in the health system.
5. How are these strategies supported? (implementation support for mechanism)	Implementation support for mechanisms is lacking i.e. refresher training to help CHWs be more effective in their work is offered erratically.
6. What social, cultural, political and economic context factors are: (1) important to the actors; (2) explain observed outcomes and (3) explain the difference between planned and actual interventions? i.e. what factors affect CHWs ability to be effective in home visits?	Community women who work as CHWs require financial compensation for their time. Women are hesitant to trust the healthcare system and frequently encouraged to stay home by older relatives. Changing gender roles affect women's agency. Many NGOs are understaffed and cannot support implementation as required.
7. What does the actual programme mean to the actors? How are they responding?	Some CHWs do not do their work because they are not motivated or supported to do so.

Table 2. Original CMO configurations for the CHW domain

Context	Programme response/mechanism	Outcome
Lack of community awareness regarding safe motherhood, danger signs during pregnancy, labour, and postpartum periods and where to go for services.	Training and deploying trusted members of the community instills 'knowledge' of danger signs during pregnancy and labour, safe delivery practices, as well as newborn care. Regular visits from community members build relationships between women and health workers, leading to 'trust' in the health system.	More women aware of complications; increased care seeking.

UNICEF's programme (CHWs), with less emphasis on the experience of programme recipients, or the generation of programme theory.

Site selection

Prior to commencing data collection, the research team reviewed existing outcome data to help identify where interventions, or components of interventions, worked and where they did not, and further clarify CMOs. Outcome data useful in understanding the impact of MNH programming might include rates of institutional delivery, number of antenatal care and postnatal care home visits or reported morbidity among mothers and newborns. In this evaluation, however, outcome data were extremely limited. Monitoring data were available at the district level only, thus failing to capture two-thirds of deliveries occurring at home (National Institute of Population and Research Training (NIPORT) *et al.* 2013), and other outcomes at the sub-district level where programme implementation occurred.

Given the lack of outcome data, qualitative indicators based on perceived programme performance were identified with UNICEF (functional referral systems, observed absenteeism, active local level planning and community engagement), and used to stratify programme areas into high and low performing. From this list, six sites were purposively selected where programme implementation had been ongoing for a year or more: one high performing district and one low performing district for each MNH programme. Site selection also considered challenging geographic areas (flood prone zones and highlands with low population density) and

cultural contexts (marginalized populations) so that programme theories could be understood in diverse settings where they might function differently. Although imperfect, this approach facilitated the examination of linkages between programme delivery approaches and outcomes.

(4). Data collection

Although the literature on realist evaluation does not advocate for one data collection methodology over another, qualitative methods are particularly adept at exploring programme context, process, theory and fidelity, and hence are prominent in the realist evaluation toolbox. A variety of qualitative methods were employed including non-participant observations, focus group discussions and in-depth interviews across the six study domains. Tools [focus group and interview guidelines, and observation checklists] were tested and refined in the field prior to data collection, and debriefing sessions enabled regular assessment of data quality and the identification of emerging issues requiring further inquiry. In total, the team developed and refined 14 different study instruments applied to a range of respondents from rural women, to CHWs, hospital staff, UNICEF field staff, implementing NGOs and government officials. In exploring the CHW domain, data collection activities included interviews with CHWs and supervisors as well as observations of home visits, community education activities, CHW training activities and interactions between CHWs and women in facilities.

(5). Data analysis

Analysis occurred in three phases. First, the research team conducted an assessment of 'programme fidelity' which involved surfacing inconsistencies between programmes as conceptualized on paper, as they existed in the minds of implementers, and what was really happening in the field. This process was facilitated by placing programmes on a 'paper to process to performance' implementation scale (Fixsen *et al.* 2009). For example, on paper, all programmes proposed a model featuring supportive and informative interactions between CHWs and women resulting in improved care seeking, and ultimately, improved health outcomes. In practice, deviations from these assumptions were apparent: limited facilitator skills, varied levels of CHW motivation and inconsistent pay and supervision resulted in failed opportunities to engage community as intended.

This initial assessment of programme fidelity was not intended to be judgmental; when programme theory is insensitive to context, programme models and their planned implementation will require adaptation to produce positive outcomes. For instance, the presence of other programmes in a shared context may necessitate unanticipated shifts in implementation strategy or focus in order to reduce duplication and improve coordination. Though not typically part of realist evaluation, the assessment of fidelity continued throughout the research period, and yielded many practical insights on how the realities of context must be taken into account to ensure successful implementation.

The second phase of analysis involved the assessment of patterns and themes using a framework analysis approach (Ritchie and Spencer 1994). Audio recordings of focus group discussions and KIIs were transcribed in Bengali and then translated into English. Over 3000 pages of transcripts were entered into Atlas TI for analysis. A standardized codebook defining domain-specific 'a priori' codes was developed, and after establishing inter and intra coder reliability, transcripts were divided among team members for coding and analysis. Each team member produced a two-page summary highlighting essential points and key quotes for quick reference, which were reviewed to identify emerging inductive codes that were subsequently applied to the data.

After the first round of coding, the research team generated domain specific reports that assembled all relevant coded text into a single file. The research team analyzed these reports using matrices to identify key findings for each domain. The final phase of data analysis involved the review of cross cutting and inductive code reports to see how they fit into domain-specific findings and whether patterns emerged across domains.

Results from analysis permitted a refinement of CMO configurations that were initially articulated based on secondary data. This refinement ensured that CMO statements reflected how the intervention functioned in broader contexts and was able to trigger (or not) mechanisms for change—thus testing the underlying assumption (programme theory), and explaining through CMO configurations what happened. Evidence of differences or adaptations to programme approaches due to context were identified and examined in terms of how they were working (or not) to achieve programme outcomes. Implementation drivers and barriers, which explained why implementation succeeded or struggled, were also explored (see Table 3). For example, evidence suggested that while CHWs may have increased community awareness, without appropriate supervision, training and support, their ability to successfully link women to facilities for EmOC was compromised. Additional barriers to facility access and care included fractured referral systems and human resource shortages in facilities. Unless UNICEF considered these barriers and/or provided options for alternatives, women would continue delivering at home.

(6). Data synthesis and reformulation of programme theories

A key phase of the realist evaluation approach is revisiting initial CMO configurations in light of evidence-based findings with the purpose of identifying recommendations that reflect the realities of implementation. The team reviewed code reports and matrices, along with original transcripts to identify supporting and contradictory data for each component of each CMO. Findings were also considered in light of whether they came from high performing or low performing sites.

As show in Table 4, reconfigured CMOs better captured the nuances of implementation, but with an emphasis on the programme vs beneficiary perspective. This focus reflected the thrust of data collection which responded to UNICEF's interest in understanding the programme functions and experience of CHWs, and receiving recommendations on how to make this cadre even more effective.

Underlying theories for each programme domain were restated to more accurately reflect what each programme was 'actually' doing and what needed to be done to generate desired changes in outcomes. This process of refining and improving specificity in CMOs based on field realities is a critical step in realist evaluation, and contributes to the design and implementation of more effective programmes.

(7). Formulation of recommendations

A final process of synthesis across domain-specific findings and rearticulated CMOs was undertaken to identify actionable recommendations. Consultative meetings between the research team and UNICEF staff ensured that recommendations were understood, relevant and feasible for UNICEF. Recommendations for CHWs went beyond the assertion that CHWs share health information, but pointed to the need for regular training, supervision and compensation to ensure that CHWs are motivated to spend sufficient time with community women and their families during home visits, and able to effect appropriate and timely referrals. Researchers realized that 'middle range theory' recommendations might be too high level for UNICEF's purposes. Although the realist evaluator may strive to create recommendations that shed light on programme theory, implementers require more process-oriented recommendations that are easily translated into action. Indeed, the research team's recommendations for CHWs were, in the end, very process-oriented given UNICEF's primary interest in identifying optimal implementation strategies across MNH programmes.

To honour the realist tradition, the research team might have resisted the implementer's urge to focus on process, and concentrated more on improvements to programme design that would better reflect programme theory. It is also the case that realist evaluation findings may result in a modification or refinement of programme theory, or a revision in the kind of intervention needed to achieve desired changes. These lessons, encountered at the end of the realist cycle, are only a few of the challenges in using realist evaluation for programme change in the context of real-time interventions.

Reflections

The steps to operationalize realist evaluation are neither straightforward nor prescribed (Greenhalgh *et al.* 2009; Marchal 2011). Although rigor was sought at every stage, and the logic of evaluation activities was maintained, the process described in this paper was iterative and sometimes difficult to navigate. In this regard, several lessons learned are offered to other novices of the approach:

Consider data quality, data volume and analysis time
High quality outcome data is a critical input to the realist approach as identifying 'what worked' requires some measurement or

Table 3. Implementation drivers and barriers for CHWs

Implementation Drivers	Implementation Barriers
1. Recruitment from the community 2. Engaged community support groups 3. Coordination with local (community) clinics 4. Standardized incentive systems 5. Organized supervision and reporting 6. Strong and ongoing training for CHWs	1. Lack of standardized salary structure 2. Limited opportunities for career advancement 3. Varying levels of training and competence to identify complications 4. Lack of supervision

Table 4. CMO configurations

Context	Mechanism	Outcome
Communities have limited access to formal health facilities/care and limited awareness of MNH care practices and danger signs. Women in communities feel isolated from health services and intimidated by health service providers.	Women in communities become CHWs, giving them an 'insight into the health system'. This 'understanding of the health system', as well as their active participation impacts their 'perception of the health facility' as well as that of the community.	CHWs are willing to encourage women to seek care in facilities—in contradiction to the prevalent practice of home delivery.
Women have few opportunities for financial and social empowerment.	Providing roles in the community for female health workers nurtures 'self-efficacy'. CHWs are 'respected' by members of their community and enjoy some 'financial independence'.	CHWs experience increased social status when they participate in programmes and enjoy financial incentives associated with their work. Their motivation supports the programme's functioning in communities.
CHWs work in rural, isolated areas with limited resources to support their work.	Providing structured systems for supervision and monitoring of CHWs makes them feel supported, 'answerable and responsible for their work'.	These systems act as implementation support to CHWs allowing messages to reach target populations. Programmes without these systems suffer; the programme model is not put into place and CHWs fail to effectively convey health messages.

assessment of outcomes. When deciding whether or not to use realist evaluation, therefore, the availability and quality of outcome data should be assessed. The kinds of outcome data a realist researcher requires may not be adequately furnished by standard systems of monitoring and evaluation, and complete and accurate information on contexts and mechanisms may be challenging to procure. Lack of adequate outcome data limited analysis, and the ability of CMOs to precisely capture the intervention process.

Data collection produced thousands of pages of text yet time for analysis was short. A multi-staged approach to managing volume was employed that relied on team-based coding and analysis. Standardized code definitions were agreed on and applied, and two-page summaries of transcripts were developed to preserve data in context. These strategies proved valuable for timely analysis, however, concerns remained that critical detail was overlooked, or patterns identified prematurely. Time was gained by splitting the team into domain-specific groups to analyze relevant reports and summaries, however, project deadlines limited opportunity for discussion and consolidation of findings, including the thorough reconfiguration of CMOs.

The analysis phase of realist evaluation is essential; indeed, it is at the heart of the approach. Ample time for analysis and iterative thinking needs to be factored into project timelines. In cases where time and resources are constrained, it is advisable to apply realist evaluation selectively to amplify understanding of more complex programme areas, and rely on routine methods for less complex domains. For example, domains such as supply side gap filling (procurement) may have been evaluated using more traditional process

evaluation approaches, while the realist evaluation of CHWs and accountability provide a nuanced understanding of systems and human elements of implementation.

Communication and commitment

Multi-disciplinary collaboration and expertise are important in the realist evaluation of public health programming, enabling insights that go beyond technical recommendations. Clear channels of communication allowed team members to share and contribute to the different phases of the research process, and to ensure comprehensive analysis. Parallel processes of data collection, analysis and consultation proved critical. Debriefing sessions in the field, and when possible, with UNICEF staff, encouraged immediate reflection on emerging findings, and identified gaps in understanding that were subsequently addressed. Without continuous engagement of this nature, the relevance of research for problem solving and improved implementation would have suffered.

Similarly, realist evaluation is contingent on the commitment of implementing agencies and donors, and assumes that findings will be used to inform how programmes are implemented. Crucial in realist evaluation is the full engagement of decision-makers at multiple levels over the course of the research process, in terms of providing feedback, and formulating and implementing recommendations for action.

Findings that implementers can act on

This realist evaluation provided recommendations that implementers could act on. Realist evaluation purists could criticize final

Table 5. Illustrative 'better' CMOs and recommendations

Illustrative CMO & Recommendation 1	In rural communities, many women distrust health facilities and health systems. This acts as a barrier to care seeking. In order to nurture trust in health systems, implementers must invest in the development of CHW cadres within the community that will be familiar, trust-worthy and accountable. Existing programme strategies such as supportive training, management and supervision help ensure that CHW cadres embody these characteristics, however, more could be done to strengthen relationships between CHW cadres and rural women. Implementers must consider how to ensure that CHWs are trust-worthy and develop flexible strategies that can respond to community women's changing expectations and needs.
Illustrative CMO & Recommendation 2	Many women expressed having limited voice in facilities. Social norms restricted women's comfort levels in requesting information from doctors or demanding any accountability at facilities. This was the case even where accountability mechanisms were in place, such as comment boxes. CHWs can fill a role in which they act as women's advocates in health centres, supporting them through delivery, and ultimately improving their delivery experience. Implementers will positively alter women's experiences in facilities when they recognize and nurture the role that CHWs can play as social supports and advocates for MNH.

recommendations for their focus on action instead of theory. For example, data clearly supported middle-range theories about the importance of trust between women and health providers, but recommendations focused on processes that would ultimately affect trust, such as improved response systems for patient complaints. Ultimately, findings identified areas of weakness in MNH programme implementation, and focused UNICEF's attention on providing greater implementation support at both facility and community levels. Arriving at useful recommendations proved to be challenging, and in retrospect, the initial conversation with UNICEF in which realist evaluation was first considered, might have better clarified its emphasis on theory generation. Had this occurred, a more traditional process evaluation approach might have been considered more aligned with their interests in identifying an 'optimal' model for MNH programming. This realization, however, was one that emerged over time and experience gained with the approach. Although CMO configurations generated through analysis were linked to process focused recommendations, the understandable interest of implementers was to identify concrete areas of intervention that would improve programme effectiveness.

Table 5 displays some illustrative CMOs and recommendations appropriate to programme implementers that better exemplify realist principles.

Conclusions

The seven steps for realist evaluation suggested in this article, emerged out of a very specific experience in using the approach. This experience also revealed the challenge of reconciling the tenets of realist evaluation, which privilege theory generation, with the more applied concerns of implementers interested in actionable recommendations on how to improve programme effectiveness. Notwithstanding this tension, a strength of the realist approach is its ability to identify and embrace the complexities of programme implementation. Realist evaluation recognizes that an intervention that creates measureable impact may depart radically from the model prescribed on paper. In some cases, this is due to the creativity and contextual sensitivity of implementers who are able to adapt programmes to effect positive change; in other cases, departure from the prescribed model is due to poor implementation practices. In short, if programme models result in undesirable outcomes, the model itself may not be at fault. Understanding the complex reasons underlying failure—whether lack of readiness for change (Weiner 2009), power dynamics or an unresponsive health system—is critical to the design of more effective interventions. Realist evaluation allows us to learn from what works and why, offering lessons on how to effect change in settings that are complex, messy, unexpected, political—and decidedly human.

References

Astbury B, Leeuw FL. 2010. Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation* **31**: 363.

Bourguignon F *et al*. 2007. Aid effectiveness—opening the black box. *The American Economic Review* **97**: 316–21.

Brownson RC *et al*. 2009. Evidence-based public health: a fundamental concept for public health practice. *Annual Review of Public Health* **30**: 175–201.

Chen HT, Rossi PH. 1983. Evaluating with sense: the theory-driven approach. *Evaluation Review* **7**: 283–302.

Dieleman M *et al*. 2009. Human resource management interventions to improve health workers' performance in low and middle income countries: a realist review. *Health Research Policy and Systems* **7**: 7.

Ekstram M. 1992. Causal explanation of social action. *Acta Sociologica* **35**: 107–22.

Fixsen DL *et al*. 2005. Implementation Research: A Synthesis of the Literature. Tampa, FL: University of South Florida, Lousi de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Greenhalgh T *et al*. 2009. How do you modernize a health service? A realist evaluation of whole-scale transformation in London. *The Milbank Quarterly* **87**: 391–416.

Harachi TW *et al*. 1999. Opening the black box: using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology* **27**: 711–31.

Hawe P *et al*. 2004. Methods for exploring implementation variation and local context within a cluster randomized community intervention trial. *Journal of Epidemiology and Community Health* **58**: 788–93.

Kazi M. 2003. Realist evaluation for practice. *British Journal of Social Work* **33**: 803–18.

Marchal B *et al*. 2010. A realist evaluation of the management of a well-performing regional hospital in Ghana. *BMC Health Service Research* **10**: 24.

Marchal B. 2011. Why do Some Hospitals Perform Better than Others? A Realist Evaluation of the Role of Health Workforce Management in Well-Performing Health Care Organisations. A Study of 4 Hospitals in Ghana. *Ph.D. Dissertation*. Vrije Universiteit Brussel, Faculty of Medicine and Pharmacy: Brussel.

Marchal B *et al*. 2012. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation* **18**: 192.

National Institute of Population and Research Training (NIPORT), Mitra and Associates and ICR International. 2013. *Bangladesh Demographic and Health Survey 2011*. Dhaka (Bangladesh) and Calverton, MA: NIPORT, Mitra and Associates, and ICF International.

Pawson R *et al*. 2005. Realist review—a new method for systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy* **10** (Suppl. 1): 21–34.

Pawson R, Manzano-Santaella A. 2012. A realist diagnostic workshop. *Evaluation* **18**: 176.

Pawson R, Tilley N. 1997. *Realistic Evaluation*. Thousand Oaks, CA: Sage.

Porter S, O'Halloran P. 2012. The use and limitation of realistic evaluation as a tool for evidence-based practice: a critical realist perspective. *Nursing Inquiry* **19**: 18–28.

Richie J, Spencer L 1994. Qualitative data analysis for applied policy research. In: Bryman and Burgess (eds), *Analysing Qualitative Data*. Routledge: London, 173–94.

Rycroft-Malone J, Fontenla M, Bick D, Seers K. 2010. A realist evaluation: the case of protocol-based care. *Implementation Science* **5**: 38.

Sanders D. 2006. Implementation research is needed to achieve international health goals. *PLoS Medicine* **3**: e186.

Sidani S, Sechrest L. 1999. Putting program theory into operation. *American Journal of Evaluation* **20**: 227.

Sullivan G *et al*. 2008. Translating clinical training into practice in complex mental health systems: toward opening the 'Black Box' of implementation. *Implementation Science* **3**: 33.

Van Belle SB *et al*. 2010. How to develop a theory-driven evaluation design? Lessons learned from an adolescent sexual and reproductive health programme in West Africa. *BMC Public Health* **10**: 741.

Weiner BJ. 2009. A theory of organizational readiness for change. *Implementation Science* **4**: 67.

Yin RK. 2003. *Case study research: design and methods*, 3rd edn. Thousand Oaks, CA: Sage.