Evaluation Research in Public Health

Angela J. Dawson

Angela J. Dawson

Australian Centre for Public and Population Health Research, University of Technology Sydney, Sydney, New South Wales

E-mail: angela.dawson@uts.edu.au

Abstract

Evaluation research is concerned with assessing the merit of health projects and programs and produces information for decision-making to improve public health. Evaluation results are critical to continuous quality improvement efforts, building organizational capacity to respond to health needs and ensuring the accountable and efficient use of resources. This chapter will introduce evaluation research to assess the outcomes of health programs and policy. The key characteristics and principles of evaluation will be examined and the range of approaches that can be taken in this applied area of research. Examples of process, outcome and impact evaluation in health contexts will enable readers to:

1. Discuss approaches to evaluation using logic models and theories of change
2. Examine program/ project evaluation designs to assess methodological rigor and appropriateness
3. Apply knowledge of global/national/state strategies and public health evidence to guide the development of evaluation indicators
4. Examine the culturally appropriate and ethically sound approaches in evaluation

Keywords: program evaluation, theory based evaluation, theory of change, logical frameworks, results based management, evaluation indicators, gender sensitive evaluation,

1. Introduction

Research and evaluation are often portrayed as a dichotomy, which is not always helpful because evaluation always employs research and, therefore, evaluations are a type of research activity with different timelines and aims. Evaluation research in public health contexts is concerned with assessing the merit of a public health project or a program and produces information for decision-making. These decisions are normally about whether the intervention or set of organized activities that comprise a program should continue to be funded modified or scaled up.

Evaluation research differs from implementation research, clinical efficacy research, and operations research. Table 1 provides an overview of the features of different approach to research including evaluation research.

Table 1. Overview of the features of different types of research

Type of research	Evaluation research	Implementation research	Translational research	Clinical efficacy research	Operations research
Characteristics					
Assess a program implementation	✓				
Assess a program effect	✓				
Identify factors that facilitate implementation effectiveness		✓			
Develop strategies to achieve effective implementation		✓			
How can evidence be applied in			✓		

practice to affect health outcomes					
Examine how a therapy works on a health outcome				✓	
Construct data based models for decision making					✓

While often the focus of evaluation research is to improve, it can also be employed to prove that the intervention is in fact responsible for change. Delivering results for and reporting to stakeholders is a feature of evaluation research that is conducted with the intent to serve the information needs of stakeholders rather than curiosity-driven research. The purpose of evaluation research is, therefore, pragmatic (Patton 2008) and is part of programmatic work often comprising twenty percent or less of the resources.

Evaluation research involves the use of both qualitative and quantitative research methods and methodologies. The study design can be descriptive or experimental while the focus can be on the effectiveness or efficiency of an intervention and/or understanding the mechanisms that help to support its implementation. According to Habicht et al. (1999, p. 11), evaluations are conducted to determine "plausibility, probability, or adequacy" of interventions. However, all evaluation research in the field of health is applied and part of a cycle of planning, implementing and assessing interventions that focus on changing people lives including the realization of their rights and improving health outcomes. This may also involve the evaluation of behavioral change and institutional change including the organization of components of health systems requiring operational change.

Learning is a key feature of evaluation research described by the European Union (2013, p. 17) as a process of learning

> through systematic enquiry what public programs and policies have achieved and understand how they perform to better design, implement and deliver future programs and policies.

2. Theory Based Evaluation

Underpinning all evaluation research is a theory or a conceptual analytical model that provides a way of structuring analysis in an evaluation. A theory is a collection of assumptions, and hypotheses that are empirically testable or that are logically connected. In the literature, theory based evaluation can be found as early as the 1930's (Coryn et al. 2011) and was further developed by key figures such as Chen (1990) and Wiess (1995). Today, theory based evaluation is commonplace and an integral part of local, national and international public health practice.

In line with an evidence-based approach to quality public health, we must ensure that our programs are underpinned and guided by principles of public health programming and that evaluation is not an ad hoc enterprise. Theory helps enhance our understanding of complex situations taking into consideration specific contextual factors. Two types of theory can be identified:

- Explanatory theory that helps to identify factors that a health program might try to change.
- Change theory that helps us to develop range of intervention strategies to address correct variables in appropriate combination with appropriate emphasis and in evaluation to assess whether all the right components are in place.

Theory, therefore, provides a meaningful way for framing or prioritizing evaluation questions. It also provides a guide to the design, and execution of the evaluation as well as the interpretation and application of the reported findings. An underpinning theory also allows programs to be generalizable to the larger population and/or transferable to other similar contexts by identifying successful elements and outcomes that can be predicted or anticipated enabling an understanding of what works and why.

A number of organizations including the expert consensus process undertaken by the Agency for Healthcare Research and Quality in the UK (Foy et al. 2011) have called for evaluation research to be integrated into the health program structure from the beginning of the planning phase to build understanding of change. This enables the team to identify which outcomes are key to the program's success and select which ones should be the focus of the evaluation.

Theory in evaluation is often driven by evaluation practice and many of the theories used have been found to been unsubstantiated by empirical studies (Coryn et al. 2011). Despite this, theory is important to the structure the planning, design and implementation of the

program and execution of the evaluation and more research is required to deliver exemplars of theory use in evaluation practice.

2.1 The Theory of change (ToC)

The theory of change (ToC) approach in evaluation is underpinned by concepts of "how and why the program will work" (Weiss 1995, p. 66) and is widespread in public health evaluations (Breuer et al. 2016). ToC as a term in evaluation emerged from social change movements and the work of the Aspen Institute on Community Change. Weiss, who was a key member of this group, decsribed the need to articulate the assumptions upon which each of the steps in a program are based in order to make the change process explicit. ToC is a causal model that explains the complexity of this change by revealing the conceptual framework that explains the causal relationships between program activities and the immediate, short term and long-term outcomes.

Evaluation theory, therefore, seeks to determine what changes have taken place at each level goal being change at many levels:

- Changes in people's lives such as the achievement of their rights and improvements in health status
- Change in the culture and organization of institutions including their values, the services they provide, legal status and their performance
- Changes in behavior such as attitudes and practices
- Change in the ways in which products and services are delivered involving improvements in knowledge and skills and cost and time effectiveness

Despite there being a lack of a definition of what at ToC is, there is agreement on the important considerations that comprise a ToC (Vogel 2012). These considerations include an explanation of the:

- Context of the initiative, i.e. the socio-cultural, political and environmental conditions, the current state of the problem the initiative is aiming to influence
- Long-term change or impact that the initiative is aiming for
- Process or stages of change expected that will lead to the desired long-term change
- Assumptions about how these changes might occur
- Outputs that are conducive to the desired change in in the specific context.
- Diagrammatic summary that outlines the change

The process of developing a ToC is usually collaborative and begins with establishing what the far-reaching outcomes or impact will be as the result of a program that are often expressed in terms of the health or social impact (see Figure 1). This is then mapped to what can be achieved in a long term such as changes in the health outcomes of a defined population and then to the immediate effects of the program upon the beneficiaries themselves. The assumptions or pre-conditions required to achieve the desired change at each stage are laid bare in a ToC including the contextual factors that may influence these necessary pre-conditions. The ToC development may also include the design of indicators to assess the change achieved through the program implementation and the evidence required to verify this.
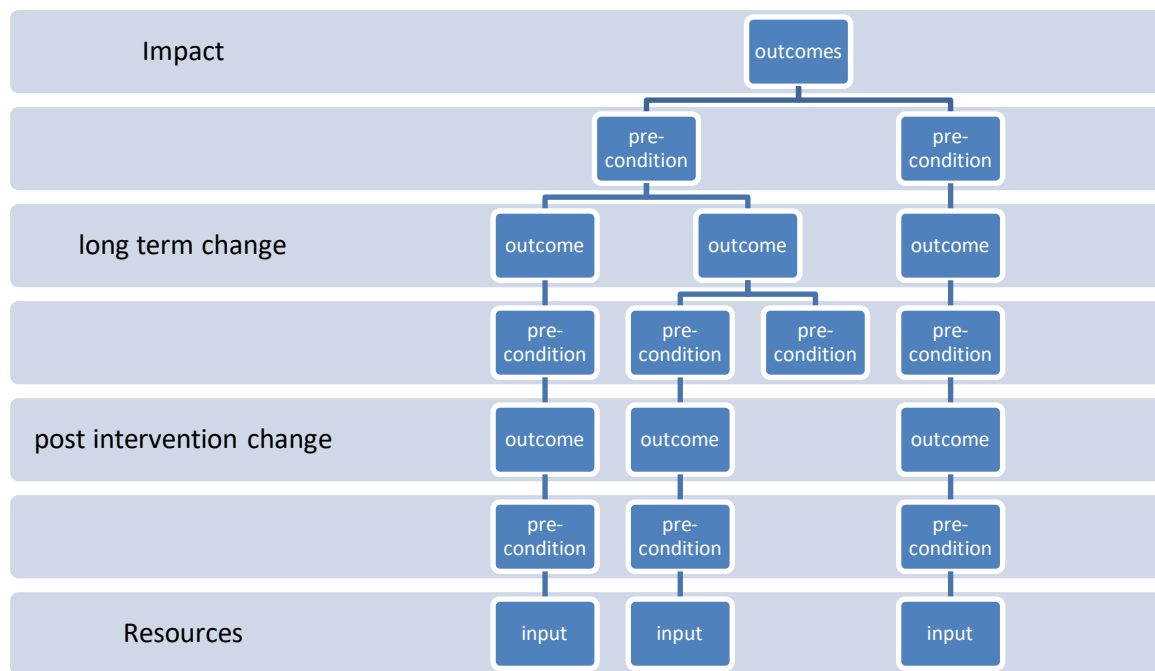

Figure 1: Theory of change (ToC) mapping

There is considerable literature to guide public health practitioners to develop their own theory of change. This includes guidance from the United Nations (Rogers 2014), philanthropic foundations (Reisman et al. 2004), universities (Taplin et al. 2013; University of Kansas 2017), community organizations (Australian Communities Foundation 2015) and networks (De Silva, Lee & Ryan 2014).

Theories of change are usually expressed graphically and in a temporal fashion from left to right. Outcomes are noted along the hypothesized causal pathway that is required to achieve the anticipated impact. There are a number of examples in the literature of these diagrams including some in the area of mental health: a Theory of Change for peer counselling for maternal depression in Goa, India (De Silva et al. 2014), the Program for Improving Mental health care cross-country summary theory of change (Breuer et al. 2015), ToC approach to develop a mental health care in a rural district in Ethiopia (Hailemariam et al. 2015) and in adolescent health (Van Belle et al. 2010; Weitzman, Silver & Dillman 2002). A worked example is provided (see Figure 2) from the community case management (CCM) project in Indonesia. CCM is a community-based service delivery model designed to address childhood illnesses such as diarrhea, pneumonia and malaria, particularly in resource poor settings (Marsh, Aakesson & Anah 2012; Setiawan et al. 2016). Here, readers can see the interventions as they pertain to political buy in, resourcing and capacity building and the effect upon treatment and care outcomes, service use, health status and costs

Political buy-in	Programme resources	Capacity building	Identification	Treatment / Care	Long- term outcome		Impact
National / Province/ District CCM approved and budget available	Medicine supply chain is functional		Essential medications for child health are available in facilities		Improved child health outcomes for those receiving care		
	Co-ordinator at post	*Continuous quality control, appropriate scope of practice, referral and supervisory mechanisms*					
Facility	health workers are available at health centres and clinics	Service providers able to diagnose and treat and refer	Health services are available, accessible, acceptable, affordable and of quality	Clearly defined package of care	Improved health worker performance	Improved child health service coverage and quality	reduced child mortality reduced health care costs
Community		sick children are identified by health workers		Sick children received required treatment and care	Improved family and child health literacy, and health seeking behaviour.		
		parents and families can identify sick children		Parents received appropriate health education	Improved compliance with care protocols and satisfaction with services		

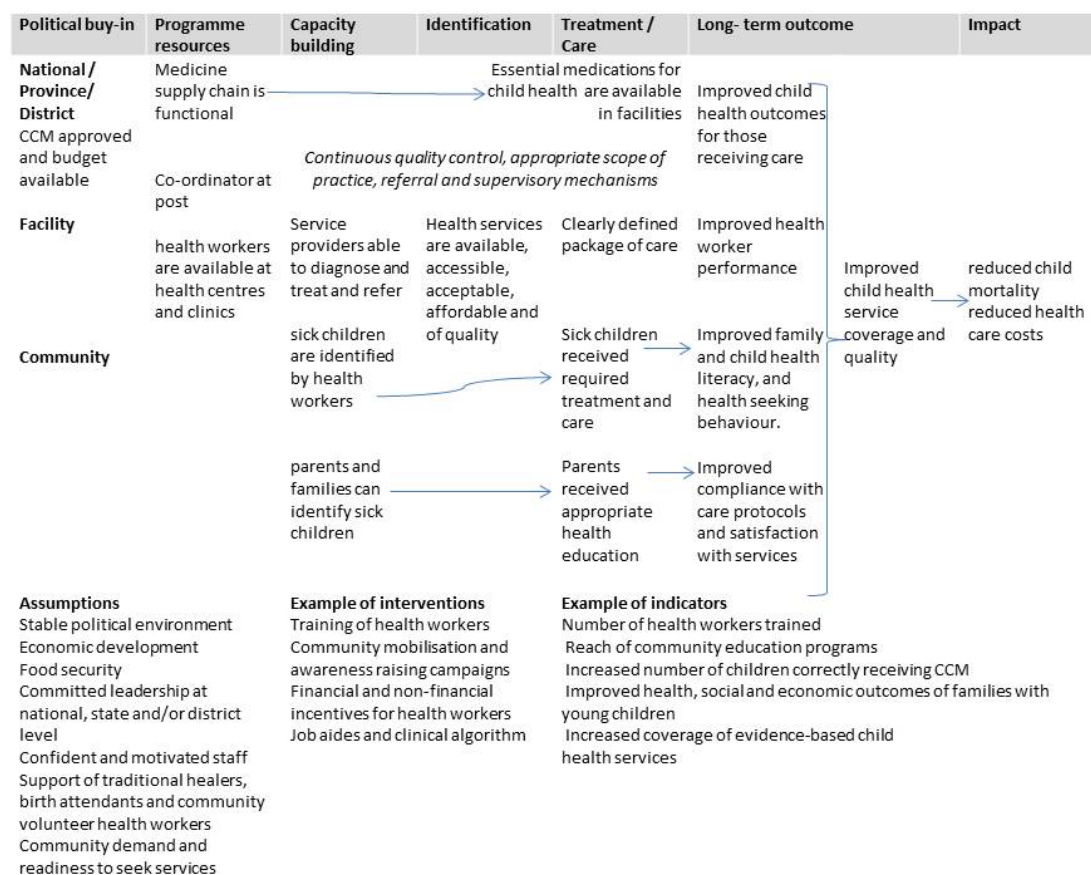| **Assumptions**	**Example of interventions**	**Example of indicators**
Stable political environment	Training of health workers	Number of health workers trained
Economic development	Community mobilisation and awareness raising campaigns	Reach of community education programs
Food security	Financial and non-financial incentives for health workers	Increased number of children correctly receiving CCM
Committed leadership at national, state and/or district level	Job aides and clinical algorithm	Improved health, social and economic outcomes of families with young children
Confident and motivated staff		Increased coverage of evidence-based child health services
Support of traditional healers, birth attendants and community volunteer health workers		
Community demand and readiness to seek services		

Figure 2: Theory of change for community case management in Indonesia

While Connell and Kubish (1998) call for credible, achievable, and testable theories of change, Breuer and colleagues (2016) have developed a useful framework that can be used to report on ToCs in public health evaluations. This consists of four elements outlined below that serve to guide those wishing to develop their own ToC.

- Clear definition of the ToC
- Description of the ToC development process (methods including stakeholder involvement)
- Summary of ToC in diagrammatic form
- Mapping of the ToC to the evaluation questions, indicators used for assessing the program's success, methods of data collection analysis and data interpretation at various time points including during and after the program implementation.

2.2 ToC and classic change theories

Theories of change, however, are not rooted in one philosophical traditional; they are pragmatic and can be strengthened by adding theories such as those from sociology or psychology. These theories can be inserted to explain change at various levels and at selected time points either before, during and after the program implementation.

Some theories focus on understanding the individual factors that influence health behavior, such as knowledge, attitudes, beliefs, and personality traits. For example, Ramsey and colleagues (Ramsay et al. 2010) have used the theory of planned behavior to examine the implementation of a knowledge translation intervention to improve the diagnostic test requesting behavior of general practitioners. A specially designed survey was used to gauge how the intervention affected the attitudes of GPs towards requesting certain tests, their beliefs about others behavior and perceptions of how easy or difficult it would be to undertake a new regime including the associated contextual factors that would hinder of facilitate this change (Ramsay et al. 2010). Other theories help to clarify processes between individuals and groups such as family, friends, peers and colleagues to explain social identity, support and roles. A post implementation evaluation of a workplace educational program to promote exercise (Amaya & Petosa 2012) used a survey based quasi-experimental design to show the effect of learning by observing others in a social context.

Evaluations that examine changes in communities can use theory to understand how organizational factors such as rules regulations and policies affect health or the effect of social norms and networks. The diffusion of innovations theory has been applied in an

evaluation of the dissemination of best practice guidelines in substance abuse treatment. The evaluation mapped the effect and rate of the uptake of the guidelines through social networks on health professional knowledge and awareness of the guidelines, how persuaded they were to change their practice, decisions taken towards change and implementation of the guidelines in services (Hubbard & Hayashi 2003).

Finally, ecological theories attempt to understand the multiple levels and see change in health behaviors, care, services and policy in terms of a complex system of interrelated factors. The California Healthy Cities evaluation framework sought to measure change at five levels: individual, civic participation, organizational, inter-organizational, and community (Kegler, Twiss & Look 2000). Bauer (1999) employed an ecological model of community organizing to evaluate a capacity and advocacy initiative for residents to impact on public health policy and training of public health professionals.

2.3 Realistic evaluation theory

Realistic evaluation is concerned with an examination of the underlying mechanisms and contextual factors that trigger change (Pawson & Tilley 1997). Many evaluation studies have developed a model of change based upon realistic theory to explain what aspects of the intervention bring about change, the extent of this and the associated contextual circumstances. In Australia, Schierhout et al.'s (2013) evaluation of a continuous quality improvement process in Indigenous health services was able to identify what worked from whom and in what contexts. Similarly Byng et al.'s (2008) evaluation of a multifaceted intervention to improve the care of people with long-term mental illness was able to develop a context-specific, mechanism-based explanations for health care effectiveness. Realistic evaluation is an iterative process that gradually reveals patterns of outcomes to determine how the program works rather than a focus on what worked.

3. Frameworks to Guide Evaluation Research

3.1 Logic models and results based frameworks

Logic models and the more a detailed form known as the logical framework or the logical framework approach (LFA) are tools designed to plan and evaluate programs and describe the goals and resources of an initiative or organization. These tools give less attention to the complex political, socio-cultural, economic, and organizational processes that underpin change in health and health care, rather they focus on the implementation of a program. LFAs

are useful to plan evaluations and employed as a metric to understand the aims, plan methods and indicators for measurement. Theory can be added to strengthen the explanation. Figure 3 lists the logic levels alongside examples of evaluation questions, the indicators employed to measure success, the means through which these indicators are verified and the underpinning assumptions upon which this change is based.

Logic model	Logic framework evaluation questions	Performance Indicators	Means of Verification	Assumptions
Impact	**Goal** To what extent have unplanned pregnancies been reduced?	Measures of goal Achievement used for evaluation	Various sources of information; methods used	Assumptions concerning Goal-purpose linkages
Outcome	**Purpose/overall objective** What increase is there in the use of family planning (FP)?	End-of-project status – to assess purpose achieved Used for project completion and evaluation	Various sources of information; methods used	Assumptions concerning the purpose/goal linkage.
	Component objectives 1. How has knowledge of FP been increased? Is there an increase in the acceptance of FP services? 2. Has the quality of FP counselling and services improved?	Measures of the extent to which component objectives have been achieved for review and evaluation.	Various sources of information; methods used	Assumptions concerning the component objective/purpose linkage.
Output	**Output /results** 1. Increased availability of educational materials 2. Improved FP supervisory system	Measures of the quantity and quality of outputs and the timing of their delivery. Used for monitoring and review.	Various sources of information; Methods used	Assumptions concerning the output/component objective linkage
Process	**Activities** 1. Community mobilization activities 2. Mass media campaign 3. Train health workers 4. Quality improvement process	Implementation/work program targets. Used during monitoring.	Project data, other sources of information	Assumptions concerning the activity/output linkage.
Input	**Inputs/ Resources** Money staff, time, political support			

Figure 3: Logic framework for the evaluation of a community based family planning program

Spearheaded by USAID, the logic framework (LF) was adopted by many donor agencies and applied across international health settings. In the late 1990s, the UN system adopted the results based management (RBM) approach in its major agencies. RBM evaluation grew out of the logical framework approach and is a management strategy that focuses on defining results based on appropriate analyses, monitoring progress, identifying and managing risks, capturing lessons learned and reporting on results achieved and resources involved. The WHO now employs a results framework to monitor the implementation of the organization's program budget, activities and outputs against its performance according to the achievement of the sustainable development goals (WHO

2017). This approach identifies the monitoring and results based evaluation phases as well as the responsibility of the WHO Secretariat and member states and partners for accountability and results.
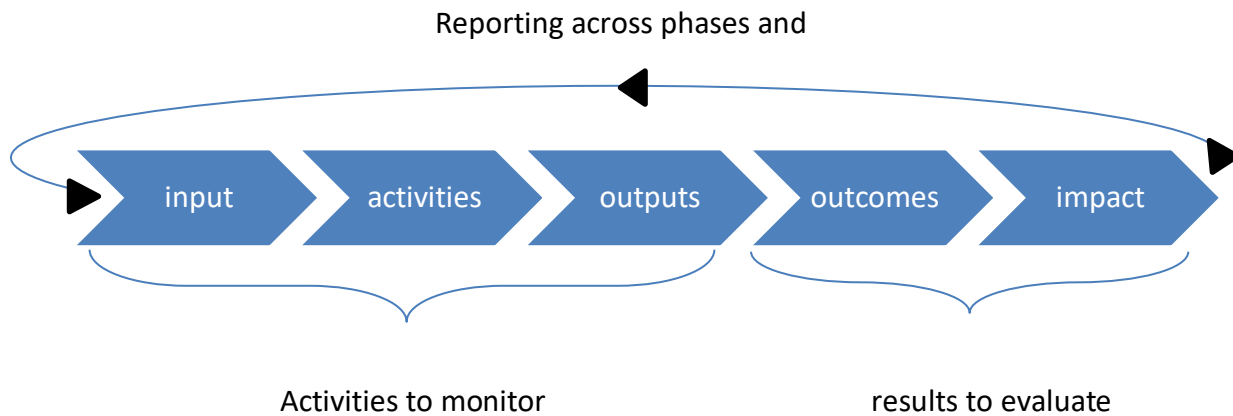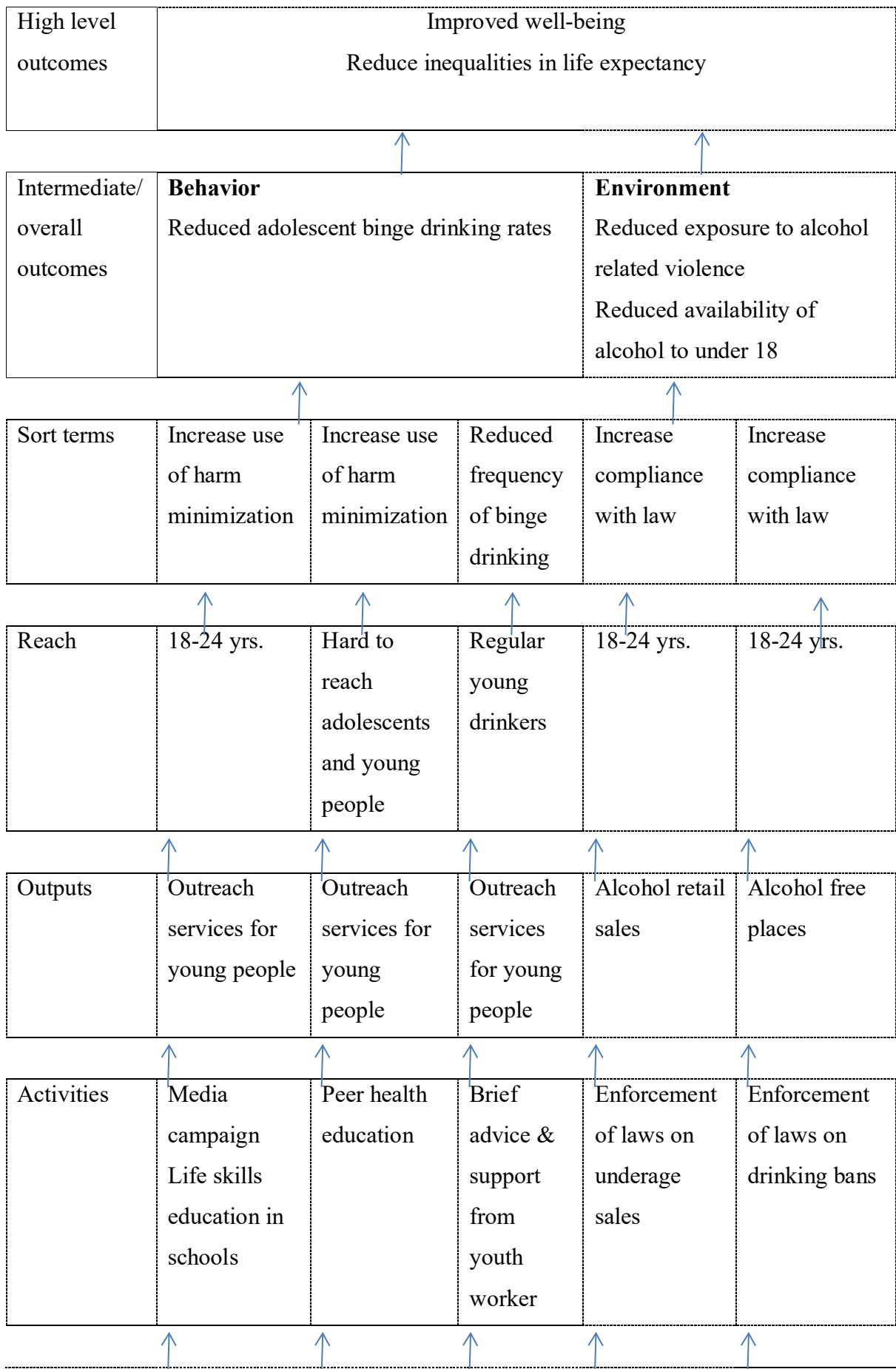
Reporting across phases and

| input | activities | outputs | outcomes | impact |

Activities to monitor   results to evaluate

Fig. 4 Results based management approach to evaluation research

While the diagram at Figure 4 represents one chain, programs are made of multiple chains that require evaluation. RBM is composed of a series of results chains (see Figure 5) that, like a logic model, is a simplified picture of an intervention designed in response to a health issue or problem and articulates the logical relationships between the resources invested, the activities, and the stages of changes that result, also known as impact.

High level outcomes	Improved well-being Reduce inequalities in life expectancy				
Intermediate/ overall outcomes	**Behavior** Reduced adolescent binge drinking rates			**Environment** Reduced exposure to alcohol related violence Reduced availability of alcohol to under 18	
Sort terms	Increase use of harm minimization	Increase use of harm minimization	Reduced frequency of binge drinking	Increase compliance with law	Increase compliance with law
Reach	18-24 yrs.	Hard to reach adolescents and young people	Regular young drinkers	18-24 yrs.	18-24 yrs.
Outputs	Outreach services for young people	Outreach services for young people	Outreach services for young people	Alcohol retail sales	Alcohol free places
Activities	Media campaign Life skills education in schools	Peer health education	Brief advice & support from youth worker	Enforcement of laws on underage sales	Enforcement of laws on drinking bans

	Schools, Govt. health promotion unit	NGOs	Govt and NGO services	Local authorities	Local authorities

Figure 5: Results chains for an evaluation of an adolescent alcohol program

The logic model approach and RBM has been criticized for being too focused on a top down and linear approach that minimizes the characteristics and expertise of people and the interaction of contextual factors on change. However, the strength of this approach is the articulation of the causal connections between conditions that need to change to reach the impact goal. A theory of change can express the assumptions that underpin the results framework.

3.2 Other frameworks to guide evaluation

In the literature, there are many other conceptual models and frameworks that can guide evaluation. The PRECEDE-PROCEED (Predisposing, Reinforcing and Enabling Constructs in Educational Diagnosis and Evaluation-Policy, Regulatory, and Organizational Constructs in Educational and Environmental Development) model was designed for health promotion planning and evaluation (Green & Kreuter 2015) has been employed in many public health interventions to evaluate workplace interventions (Post et al. 2015) to individual chronic disease programs (Azar et al. 2017). The Re-Aim (Reach, Effectiveness, Adoption, Implementation, Maintenance) framework (Glasgow, Vogt & Boles 1999) is another useful tool to structure evaluations of individual (Belkora et al. 2015) and community (Jenkinson, Naughton & Benson 2012) and partnership (Sweet et al. 2014) initiatives.

The Centers for Disease Control and Prevention in the United States has developed a Framework for Program Evaluation in Public Health (CDC 1999). This framework summarizes the key elements of evaluation and proposes a six-stage cycle comprised of engaging stakeholders, articulating the program and the evaluation design, gathering credible evidence, justifying the conclusions reached and sharing lessons learned. This is coupled with standards for effective program evaluation that have been applied in public health disease control programs (Logan et al. 2003).

More recent frameworks include Proctor et al.'s (2011) eight conceptually distinct outcomes for potential evaluation: acceptability, adoption (also referred to as uptake), appropriateness, costs, feasibility, fidelity, penetration (integration of a practice within a

specific setting), and sustainability (also referred to as maintenance or institutionalization). This has been largely applied in implementation research such as the population based care program for those at risk for delirium, alcohol withdrawal, and suicide harm (Lakatos et al. 2015). Finally, another potentially useful approach to evaluation design is ten steps to making evaluation matter outlined by Sridharan and Nakaima (2011) that add considerations from the realist tradition including sustainability, and learning considerations.

4. Purpose and Phases of Evaluation Research

Evaluation may be shaped by the purpose for which it is designed as well as the time-frame in which it is executed. As Habicht et al. (1999) suggest, the purpose of an evaluation research can be to establish plausibility, adequacy or probability. If the aim is plausibility, then the focus will be on designing the evaluation to reveal best how a program achieved its expected objectives and that the change that occurred during the process can potentially be attributed to the program activities. If the aim is to determine adequacy, then this will be an evaluation that seeks to establish if the program goals were achieved. However, an evaluation with the goal of determining probability will most likely employ an experimental design to demonstrate that improved health outcomes or impact is directly attributed to the program activities.

In addition to this, there are several phases or stages of a program implementation where evaluation research can be undertaken as outlined in Figure 6. This can proceed the design and implementation of a health program or intervention so that baseline data can be collected to not only inform the design of the intervention, but also provide a yardstick for measuring change. The next phase of evaluation might involve piloting or testing aspects of the intervention to ensure feasibility, appropriateness or fit. This process may involve some modification of the intervention and provide additional baseline data. Implementation or process evaluation is known as "real time" evaluation and involves the regular collection and reporting of information to track whether activities are being implemented and immediate results are achieved as planned (Moore et al. 2015). Theory can be useful to structure this evaluation (Ramsay et al. 2010). Post implementation reviews stake place immediately after rather than during the implementation of an intervention while outcome evaluation and impact evaluation map short term and longer-term change respectively. Outcome and impact evaluation are often termed summative evaluation and aim to answer specific questions about performance of the activities. They are concerned with answering how and why questions linked to plausibility or causality.

Time ————————————————————————————————————→

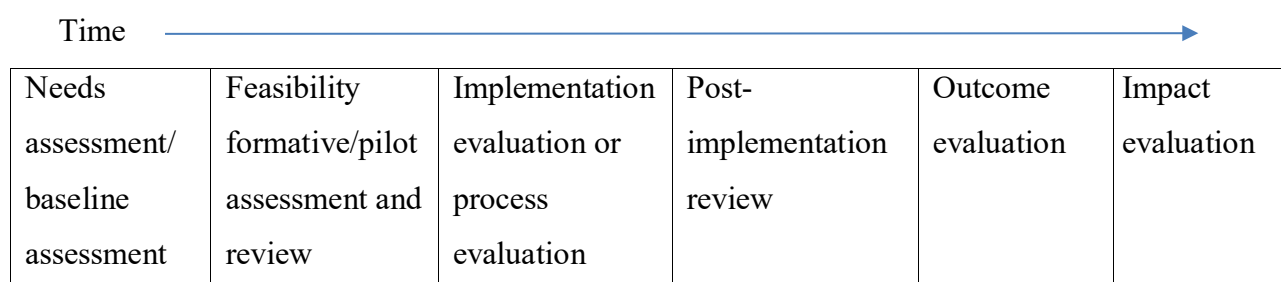| Needs assessment/ baseline assessment	Feasibility formative/pilot assessment and review	Implementation evaluation or process evaluation	Post-implementation review	Outcome evaluation	Impact evaluation

Figure 6: Phases or stages where evaluation research can be undertaken

There are a number of useful guides to these various types of evaluations in public health contexts provided by departments of Health (ACI 2013), international non-government organizations (IFRC 2011) and the United Nations (UNDP 2009; WHO 2013).

Different types of evaluations may be undertaken across these phases that draw on both qualitative and qualitative evidence. Pre-intervention evaluations may comprise: assessments of health needs that involve surveys of or interviews with community members, or from existing statistical health data; desk reviews of existing reports and policy documentation; or financial audit and risk assessments of the context into which a program or policy may be implemented. Economic evaluations including cost effectiveness assessment and cost-benefit analysis can be undertaken across all phases alongside quantitative analysis and qualitative evaluations involving observations of behavior, key informant interventions and participatory processes.

5. Evaluation Research Designs

Selecting the study design for an evaluation depends on the purpose of the evaluation. This will dictate the stage or phase where it is carried out and the type of evaluation. For example, an evaluation whose purpose includes is to understand whether the budget was allocated effectively or health staff performance during the implementation of a program may involve systems to monitor the finances or standards over a specific time-frame. Other evaluation activities might involve an examination of changes in knowledge or behaviors such as the uptake of contraception. These activities could be part of a process evaluation and employ a quasi-experimental pre-and post-intervention design. Such activities contrast with experimental longitudinal designs where causal links are sought to identify if the program demonstrated an impact on health outcomes of the beneficiaries or the larger population. Impact and outcome evaluation may also involve mixed methods combining for example

ethnography involving the data collection from in-depth interviews and observation with survey and/or population based surveillance data. Table 2 identifies some characteristics and examples of experimental, qualitative and mixed methods evaluation designs in public health. However, it is possible that an evaluation of a program could be comprised of all or some of these designs and methodologies.

Table 2: Study designs and methodologies for evaluation research

Study design	Explanation and example
Experimental and quasi-experimental evaluation designs	
Randomized control trial	The health program's impact is the outcome of interest. Common form involves one group being randomly assigned to receive the intervention and the other receives no intervention or usual treatment (**see also Randomized Controlled Trials**). Useful when intervention is introduced in small population in highly structured manner, see in the case of the evaluation of a mindfulness program (Hou et al. 2014). Limited by high resource implications and does not necessarily reflect how interventions will work beyond the experiment.
Quasi-experimental, comparison group design.	May involve a study of a group before and after receiving an intervention. A comparison group could be included. See an example in the evaluation of an urban health initiatives (Weitzman, Silver & Dillman 2002)
Economic evaluation	Statistical measurement of the inputs and outcomes of an evaluation to examine the costs and consequences of an initiative. Sinha et al. (2017) undertook a cost benefit analysis of a program involving women's groups facilitated by community workers to reduce neonatal mortality in rural India.
Qualitative evaluation	
Ethnography	This methodology involves the study of culture using observation, in-depth interview and field notes. It involves the researcher spending long periods in the field studying knowledge systems of

	groups of people (**see Critical Ethnography in Public Health, Ethnographic Method & Institutional Ethnography**). Ethnography has been applied in the formative evaluation of infant feeding initiatives (Young & Tuthill 2017).
Mixed methods	
Participatory evaluation	An approach that engages stakeholders in design, planning and undertaking the evaluation with the goal of improving skills and ensuring more responsive health care and services (**see also Community-Based Participatory Action Research**). One example from mental health involve consultation with consumers, community people and providers to contextualize and validate the findings from case studies (Lea et al. 2015).
Realist evaluation	Theory driven evaluation to determine the contextual mechanisms that enable the successful achievement of program outcomes. Qualitative and quantitative approaches are employed according to what best answers the questions. Pragmatic design visible in an evaluation of continuous quality improvement in primary health care (Schierhout et al. 2013)
Developmental evaluation	An approach that is responsive to context by allowing constant adaption and enables the gathering of real time data. Suits complex situations for example the evaluation of social change in communities (Patton, McKegg & Wehipeihana 2016).

6. Evaluation Indicators

An indicator is a variable that provides accurate and reliable evidence about the achievement of a specific result. Indicators should be observable, well-defined, measurable, and agreed upon. They can be both qualitative and quantitative and are at all levels of the program logic or results chain. Indicators that make up a process evaluation usually involve the regular collection and reporting of data to monitor whether results are being realized as planned and to identify problem areas and possible solutions. Such indicators are often found in processes of continuous quality improvement efforts and require an operational definition.

Indicators focused on assessing the achievement of results in outcome and impact evaluations are analytical efforts to answer specific questions about performance of program activities. There are generally concerned with answering questions concerning why the intended outcomes were or were not realized and how the results were achieved. Such indicators are designed to determine the probability of a program to health and social outcomes over time or the causal contributions of activities to results to confirm a hypothesis.

In Table 2, we can see that the evaluation questions outlined in Figure 3 have been formed into objectives that have been further qualified by indicators across the various evaluation levels. These indicators relate to the provision, utilization, coverage and impact of health services as well as the legal and social environment. Other indicators can include:

- Improved health outcomes
- Increased use of health facilities
- Extension of quality health services
- Development of human resources for health
- Improved legal environment
- Achieve gender equality

Other indicators could include:

- Improved economic productivity
- Improved social capital that includes the use of social networks to improve health this includes the facilitation of co-operation and mutually supportive relations in communities to reduce social isolation, improve well-being and harness the skills and talents of individual, increase access to employment and education opportunities
- Improved cultural capital education (knowledge and skills) that provides advantage in achieving a higher social-status in society.

Table 3: Examples of evaluation objectives and indictors at impact, outcome and impact levels

Impact Evaluation	Impact evaluation indicator
Impact objective Reduce adolescent fertility	Adolescent birth rate (aged 10-14 years; aged 15-19 years) per 1,000 women in that

	age group reduced by three quarters in country x by 2030
Outcome evaluation	**Outcome evaluation indicator**
Overall outcome evaluation objective 1 Increase adolescent use of modern methods of contraception	Contraceptive prevalence rate in province X increased by x
Overall outcome evaluation objective 2 Improved social and policy environment for contraception and sexual and reproductive health and rights	Institution of laws and regulations that guarantee women aged 15-19 years access to sexual and reproductive health care, information and education
Component 1 outcome evaluation objective Increased uptake of adolescent contraception services	% of new clients and return of clients
Component 2 outcome evaluation objective Improved quality of contraceptive counseling and services for adolescents	% of sites adhering to adolescent friendly standards
Component 3 outcome evaluation objective Increased access to contraception services	% satisfaction
Component 4 outcome evaluation objective Increase availability of contraceptive commodities	% of functional procurement and distribution in the supply chain
Component 5 outcome evaluation objective Increase in female adolescent reproductive health decision-making	Proportion of female adolescents who make their own informed decisions regarding sexual relationships, contraceptive use and reproductive health care
Output evaluation	
Component 1 output evaluation objective Increased adolescent knowledge and acceptance of modern methods of contraception and service location	% of adolescents with knowledge of available services and commodities

	Positive attitudes towards contraception and increased expressed demand
Component 2 output evaluation objective Improved health workers contraceptive counseling skills	% of staff trained and assessed as competent
Component 3 output evaluation objective Appropriate clinic opening hours, timeliness of consultation and appropriate staffing numbers	% of facilities with minimum staffing norms (List of minimum staffing defined)
Component 4 output evaluation objective Health centers stocked with low cost essential RH commodities	% of facilities without 7-day stock outs of essential drugs (List of essential drugs defined)
Component 5 output evaluation objective Increase in contraception services at health clinics	% of services delivering evidence based contraceptive services, care and information to adolescents

There is considerable generic guidance on developing quality indicators for evaluation in general they should be valid, reliable, precise, timely and comparable. Table 4 defines these attributes using indicators from a family planning evaluation as an example.

Table 4: Attributes of quality indicators

Attribute	Example of indicator
Valid	Participants will recall/describe at least three modern methods of family planning
Reliable	The indicator above could be used and classified as reliable if in pre-testing different people (interviewer and participants) demonstrated a consistent understanding of the term "modern". If not, then validity may be affected since different people may understand different methods as modern.
Precise	The indicator must be able to be clearly defined. In this case a pre-defined list of modern family planning methods should be able to be

	produced. The indicator must be precise so that the answers can be clearly assessed.
Timely	Change in this indicator could be expected to be within a short time frame. However, if the evaluation sought to measure change in family size in a 2–3-year project it will not be possible to observe such an indicator within the time-frame of the project.
Comparable	Knowledge of three modern family planning methods should be comparable across various populations. It should be straightforward to make a comparison between men's and women's knowledge of contraception. However, if an intervention-specific indicator was selected, for example if we wanted to know how many modern methods that adolescents who are peer health educators can list, this is only useful for that group of people but could not be applied to other groups.

7. Considerations in the Development of Indicators

One of the issues evaluation research in is ensuring that everyone involved is applying the same assessment framework to the measurement of outcomes. An operational definition of each indicator is, therefore, required so that those involved in collecting data can assess the achievement of the indicator in a standard manner. This also requires that the evaluation design is rigorous and aligned with best practice efforts that provide comparable data on changes over time. A protocol is also required to guide data collection, as well as standard tools to collect such data. Piloting or testing indicators in the field with the proposed data gathering tools is useful to ensure that all issues can be addressed before the roll out.

Another area to consider when developing indicators is how they might best connect with existing measures and could be integrated across the health system to provide a useful picture of change. Indicator designs can, therefore, benefit from being aligned with global national/state strategies and public health evidence. This enables comparability and although they may need to be field tested for the unique context of your evaluation, they will already be quite sturdy. For example, countries may already have goals and measures by which they would like to reduce the adolescent fertility rate in line with their Sustainable Development Goals (SDG) targets and measures. This indicator could be inserted at table 2 to specific the impact evaluation objective and indicator. Other SDG target and goals may also be relevant

here such as existing country indicators to achieve gender equality and empower all women and girls.

Some indicators such as the measurement of community participation will require extensive consultation to ensure that what is measured is appropriate and sound. For example, several indicators maybe required to evaluate community engagement in a participatory action and learning health initiative. The list below outlines many indicators that could be included in an evaluation.

- No. and % of activities that had a record of community participation
- No. and % activities where community members were involved in identifying the problem or issue
- No. and % activities where community members were involved in determining strategies (deciding what to do) about the problem or issue
- No. and % activities where community members were involved in implementing the strategy (doing the work)
- No. and %. activities where community members were involved in evaluating the results of the work

However, the measures of these indicators will be dependent on the capacity of the community to participate including the skills and knowledge of the people, the strength of the community organizations and stability of the political and economic context. It is necessary in an evaluation to have buy in from all sectors, particularly the community to ensure success.

8. Culturally Appropriate, Gender Sensitive, Ethically Sound Evaluation

Engaging stakeholders including health professionals, decision-makers and community members before, during and after evaluation research is essential to ensure that the evaluation questions and indicators are relevant and appropriate, and that data is ethically collected. It is critical to include sex, gender, culture/ethnicity and age categories for data collection as this helps to identify norms, values, attitudes and behaviors that may affect health and the impact of a program. Gender norms, for example, can be a basis for discrimination and bias. Gender norms around early marriage can work to a girl's disadvantage by preventing their engagement in education, fulfilling employment and predispose them to early childbearing and associated death and disability.

While sex-disaggregated data (data that are collected, analyzed, and reported for men and women and boys and girls separately) is useful, gender-sensitive indicators can be effective in measuring gender or social differences between the sexes. These indicators can measure changes in status, roles, expectations, and norms pertaining to people based on what gender they are or identify themselves as. Gender-sensitive indicators vary in complexity, with some requiring more elaborate data collection or analytic methods than others. For example, the proportion of people (disaggregated by sex) who can make decisions about their own health care/health care for their children, or the proportion of people (disaggregated by sex) who experienced physical violence from an intimate partner in the last 12 months. As many of these indicators require the collection of sensitive data, consent and ethical processes are mandatory as is the case will all evaluation research where the results are to be published. However as many evaluations are internal processes and ethical approval may not be required. The collection of data against gender sensitive indicators may also require the employment of field workers of the same gender, culture and religion to ensure that participants are comfortable in responding.

Effectively engaging stakeholders as equal partners facilitates ownership over the evaluation process and outcomes to ensure that modifications to the program are made during implementation evaluation and lessons transferred in policy and practice.

Thought needs to be given to who should be involved and how this might contribute to the effect of the actual intervention. For example, engaging men in discussions about how the outcomes of a maternal health program might be evaluated or how the results can be applied may increase husband's participation in birth preparedness, a known factor to improve maternal health outcomes. Involving men may also facilitate women's access to facilities in cases where men's approval must be given, and finances maybe required to travel to a health clinic. Training and involving community midwives in collecting data as part of a maternal health evaluation at village level may provide the most up-to-date information on women who are pregnant in rural situations where data collection is poor.

9. Conclusion and Future Directions

The goal of evaluation research is utilization in policy and practice to improve quality of life. A balance must be, therefore, achieved between quality data and rigorous processes and ensuring that there is ownership and involvement of all stakeholders so that change and health improvement can be actioned.

In the end, the measure of our success will not be predicated on the number of evaluations done, or stored within a database, or even solely upon the quality of the findings.…Our success will depend on our ability to use evaluation findings to strengthen our efforts and sharpen our decision-making." (USAID 2011, pp. Rajiv Shah, Administrator, Preface)

Success in evaluation is not always communicated past the reports to funders due to budget and time constraints. However, while sharing lessons learned in peer-reviewed literature is important, so too is the dissemination of evaluation results in the form of practice or policy options briefs for decision-makers. Such dissemination formats help to make evaluation findings accessible and organizations accountable for the resources used. Documenting and sharing evaluation knowledge is, therefore, key to institutionalizing health improvement efforts.

Institutionalizing data-informed decision making derived from evaluation research is likely to become a key part of future practice with technology playing a central role. Instead of establishing systems to collect data evaluators are likely to become more involved in data mining and data linkage activities using existing sources that will enable real time evaluation across multiple sites and countries. The internet may accommodate an increased and participatory approach to evaluation research where citizens and stakeholders can offer comments, contribute data and undertake analyses. This will facilitate evaluations that capture and respond to the socio-cultural diversity in society locally and globally. Evaluation research processes are also likely to become more transparent with activities taking place in on-line open access platforms that enable learning to be easily accessed and shared. With these changes may come challenges that could affect the independent nature and quality of evaluation research that standard education and the professionalization of the field can help to keep in check.

References

ACI 2013, *Understanding Program Evaluation An ACI Framework*, Agency for Clinical Innovation, NSW Department of Health, Chatswood NSW Australia.

Amaya, M. & Petosa, R. 2012, 'An evaluation of a worksite exercise intervention using the social cognitive theory: A pilot study', *Health Education Journal*, vol. 71, no. 2, pp. 133-43.

Australian Communities Foundation 2015, *Theory of Change*, Australian Communities Foundation, Fitzroy, Victoria, viewed 17 January 2018, <http://www.communityfoundation.org.au/about-acf/theory-of-change/>.

Azar, F.E., Solhi, M., Nejhaddadgar, N. & Amani, F. 2017, 'The effect of intervention using the PRECEDE-PROCEED model based on quality of life in diabetic patients', *Electronic physician*, vol. 9, no. 8, pp. 5024–30.

Belkora, J., Volz, S., Loth, M., Teng, A., Zarin-Pass, M., Moore, D. & Esserman, L. 2015, 'Coaching patients in the use of decision and communication aids: RE-AIM evaluation of a patient support program', *BMC health services research*, vol. 15, no. 209, pp. doi.org/10.1186/s12913-015-0872-6.

Breuer, E., De Silva, M., Shidaye, R., Petersen, I., Nakku, J., Jordans, M., Fekadu, A. & Lund, C. 2015, 'Planning and evaluating mental health services in low-and middle-income countries using theory of change', *The British journal of psychiatry*, vol. s1-s8, p. doi.org/10.1192/bjp.bp.114.153841

Breuer, E., Lee, L., De Silva, M. & Lund, C. 2016, 'Using theory of change to design and evaluate public health interventions: a systematic review', *Implementation Science*, vol. 11, no. 63, pp. doi.org/10.1186/s13012-016-0422-6.

Byng, R., Norman, I., Redfern, S. & Jones, R. 2008, 'Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation', *BMC Health Services Research*, vol. 8, no. 271, pp. doi.org/10.1186/472-6963-8-274.

CDC 1999, *Framework for Program Evaluation in Public Health* *https://www.cdc.gov/mmwr/preview/mmwrhtml/rr4811a1.htm* CDC Evaluation Working Group, Centers for Disease Control and Prevention, No.RR-11.

Chen, H.T. 1990, *Theory-driven evaluations*, Sage., Thousand Oaks, CA.

Coryn, C.L., Noakes, L.A., Westine, C.D. & Schröter, D.C. 2011, 'A systematic review of theory-driven evaluation practice from 1990 to 2009', *American journal of Evaluation*, vol. 32, no. 2, pp. 199-226.

De Silva, M., Lee, L. & Ryan, G. 2014, *Using Theory of Change in the development, implementation and evaluation of complex health interventions A practical guide*, The Centre for Global Mental Health & the Mental Health Innovation Network, London.

De Silva, M.J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C. & Patel, V. 2014, 'Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions', *Trials*, vol. 15, no. 267, pp. doi.org/10.1186/745-6215-15-267.

EU 2013, *EVALSED: The resource for the evaluation of Socio-Economic Development, http://ec.europa.eu/regional_policy/en/information/publications/evaluations-guidance-documents/2013/evalsed-the-resource-for-the-evaluation-of-socio-economic-development-evaluation-guide*, European Commission

Foy, R., Ovretveit, J., Shekelle, P.G., Pronovost, P.J., Taylor, S.L., Dy, S., Hempel, S., McDonald, K.M., Rubenstein, L.V. & Wachter, R.M. 2011, 'The role of theory in research to develop and evaluate the implementation of patient safety practices', *Quality and Safety in Health Care*, vol. 20, no. 5, pp. 453-9.

Glasgow, R.E., Vogt, T.M. & Boles, S.M. 1999, 'Evaluating the public health impact of health promotion interventions: the RE-AIM framework', *American journal of public health*, vol. 89, no. 9, pp. 1322-7.

Green, L. & Kreuter, M. 2015, *Health program planning: an educational and ecological approach*, vol. 4, McGraw-Hill Higher Education, New York.

Habicht, J.-P., Victora, C. & Vaughan, J.P. 1999, 'Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact', *International journal of epidemiology*, vol. 28, no. 1, pp. 10-8.

Hailemariam, M., Fekadu, A., Selamu, M., Alem, A., Medhin, G., Giorgis, T.W., DeSilva, M. & Breuer, E. 2015, 'Developing a mental health care plan in a low resource setting: the theory of change approach', *BMC health services research*, vol. 15, no. 429, pp. doi.org/10.1186/s12913-015-1097-4.

Hou, R.J., Wong, S.Y.-S., Yip, B.H.-K., Hung, A.T., Lo, H.H.-M., Chan, P.H., Lo, C.S., Kwok, T.C.-Y., Tang, W.K. & Mak, W.W. 2014, 'The effects of mindfulness-based stress reduction program on the mental health of family caregivers: a randomized controlled trial', *Psychotherapy and psychosomatics*, vol. 83, no. 1, pp. 45-53.

Hubbard, S.M. & Hayashi, S.W. 2003, 'Use of diffusion of innovations theory to drive a federal agency's program evaluation', *Evaluation and program planning*, vol. 26, no. 1, pp. 49-56.

IFRC 2011, *Project/programme monitoring and evaluation (M&E) guide*, International Federation of Red Cross and Red Crescent Societies, Geneva.

Jenkinson, K.A., Naughton, G. & Benson, A.C. 2012, 'The GLAMA (Girls! Lead! Achieve! Mentor! Activate!) physical activity and peer leadership intervention pilot project: a process evaluation using the RE-AIM framework', *BMC Public Health*, vol. 12, no. 1, p. 55.

Kegler, M.C., Twiss, J.M. & Look, V. 2000, 'Assessing community change at multiple levels: the genesis of an evaluation framework for the California Healthy Cities Project', *Health Education & Behavior*, vol. 27, no. 6, pp. 760-79.

Lakatos, B.E., Schaffer, A.C., Gitlin, D., Mitchell, M., Delisle, L., Etheredge, M.L., Shellman, A. & Baytos, M. 2015, 'A population-based care improvement initiative for patients at risk for delirium, alcohol withdrawal, and suicide harm', *The Joint Commission Journal on Quality and Patient Safety*, vol. 41, no. 7, pp. 291-AP3.

Lea, S., Callaghan, L., Eick, S., Heslin, M., Morgan, J., Bolt, M., Healey, A., Barrett, B., Rose, D. & Patel, A. 2015, *The management of individuals with enduring moderate to severe mental health needs: a participatory evaluation of client journeys and the interface of mental health services with the criminal justice system in Cornwall* *https://www.ncbi.nlm.nih.gov/books/NBK285789/* National Intistute of Health Research, Southampton (UK), No. 3.15.

Logan, S., Boutotte, J., Wilce, M. & Etkind, S. 2003, 'Using the CDC framework for program evaluation in public health to assess tuberculosis contact investigation programs', *The International Journal of Tuberculosis and Lung Disease*, vol. 7, no. 12, pp. S375-S83.

Marsh, D., Aakesson, A. & Anah, K. 2012, *Community Case Management Essentials: Treating Common Childhood Illnesses in the Community. A Guide for Program Managers*, CORE Group Save the Children BASICS MCHIP, Washington, D.C.

Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T. & Wight, D. 2015, 'Process evaluation of complex interventions: Medical Research Council guidance', *BMJ*, vol. 350, p. doi.org/10.1136/bmj.h258.

Patton, M.Q. 2008, *Utilization-focused evaluation*, 4 edn, Sage, Thousand Oaks, CA.

Patton, M.Q., McKegg, K. & Wehipeihana, N. 2016, *Developmental Evaluation Exemplars: Principles in Practice*, Guilford Press, New York.

Pawson, R. & Tilley, N. 1997, *Realistic Evaluation*, Sage, London.

Post, D.K., Daniel, M., Misan, G. & Haren, M.T. 2015, 'A workplace health promotion application of the Precede-Proceed model in a regional and remote mining company

in Whyalla, South Australia', *International Journal of Workplace Health Management*, vol. 8, no. 3, pp. 154-74.

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', *Administration and Policy in Mental Health and Mental Health Services Research*, vol. 38, no. 2, pp. 65-76.

Ramsay, C.R., Thomas, R.E., Croal, B.L., Grimshaw, J.M. & Eccles, M.P. 2010, 'Using the theory of planned behaviour as a process evaluation tool in randomised trials of knowledge translation strategies: a case study from UK primary care', *Implementation Science*, vol. 5, no. 71, pp. doi.org/10.1186/748-5908-5-71.

Reisman, J., Gienapp, A., Langley, K. & Stachowiak, S. 2004, *Theory of Change A Practical Tool for Action, Results and Learning http://www.aecf.org/resources/theory-of-change/* Organizational Research Services, Annie E. Casey Foundation Seattle Washington.

Rogers, P. 2014, *Theory of Change, Methodological Briefs: Impact Evaluation 2*, UNICEF Office of Research, Florence.

Schierhout, G., Hains, J., Si, D., Kennedy, C., Cox, R., Kwedza, R., O'Donoghue, L., Fittock, M., Brands, J. & Lonergan, K. 2013, 'Evaluating the effectiveness of a multifaceted, multilevel continuous quality improvement program in primary health care: developing a realist theory of change', *Implementation Science*, vol. 8, no. 119, pp. doi.org/10.1186/748-5908-8-119.

Setiawan, A., Dignam, D., Waters, C. & Dawson, A. 2016, 'Improving access to child health care in Indonesia through community case management', *Maternal and child health journal*, vol. 20, no. 11, pp. 2254-60.

Sridharan, S. & Nakaima, A. 2011, 'Ten steps to making evaluation matter', *Evaluation and program planning*, vol. 34, no. 2, pp. 135-46.

Sweet, S.N., Ginis, K.A.M., Estabrooks, P.A. & Latimer-Cheung, A.E. 2014, 'Operationalizing the RE-AIM framework to evaluate the impact of multi-sector partnerships', *Implementation Science*, vol. 9, no. 1, p. 74.

Taplin, D., Clark, C., Collins, E. & Colby, D. 2013, *Theory of Change A Series of Papers to Support Development of Theories of Change Based on Practice in the Field.*, Center for Human Environments, New York.

UNDP 2009, *Handbook on Planning,Monitoring and Evaluating for Development Results http://www.undp.org/eo/handbook*, United Nations Development Programme, New York, NY 10017, USA.

University of Kansas 2017, *Developing a Framework or Model of Change*, Center for Community Health and Development, University of Kansas, Kansas, viewed 17 January 2018, <http://ctb.ku.edu/en/4-developing-framework-or-model-change >.

USAID 2011, *Evaluation learning from experience: USAID evaluation policy*, U.S. Agency for International Development., Washington, DC.

Van Belle, S.B., Marchal, B., Dubourg, D. & Kegels, G. 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*, vol. 10, no. 741, pp. doi.org/10.1186/471-2458-10-741.

Vogel, I. 2012, *Review of the use of 'Theory of Change' in international development Review Report*, UK Department of International Development London.

Weiss, C.H. 1995, 'Nothing As Practical As Good Theory: Exploring Theory-Based Evaluation For Comprehensive Community Initiatives For Children And Families', in J.P. Connell, A.C. Kubisch, L. Schorr, B. & C.H. Weiss (eds), *New Approaches to*

Evaluating Community Initiatives: Concepts, Methods and Context, The Aspen Institute, Washington DC, pp. 65-92.

Weitzman, B.C., Silver, D. & Dillman, K.-N. 2002, 'Integrating a comparison group design into a theory of change evaluation: The case of the urban health initiative', *American Journal of Evaluation*, vol. 23, no. 4, pp. 371-85.

WHO 2013, *WHO evaluation practice handbook*, World Health Organization, Geneva.

WHO 2017, *WHO's results framework* World Health Organization, Geneva, viewed 21 October 2017, <http://www.who.int/about/who_reform/change_at_who/results_framework/en/#.WfLd4rVx2Uk www.who.int/about/resources_planning/WHO_GPW12_results_chain.pdf >.

Young, S.L. & Tuthill, E. 2017, 'Ethnography as a tool for formative research and evaluation in public health nutrition: illustrations from the world of infant and young child feeding', in J. Chrzan & J. Brett (eds), *Research Methods for Anthropological Studies of Food and Nutrition: Food Research*, Berghahn Books, New York.