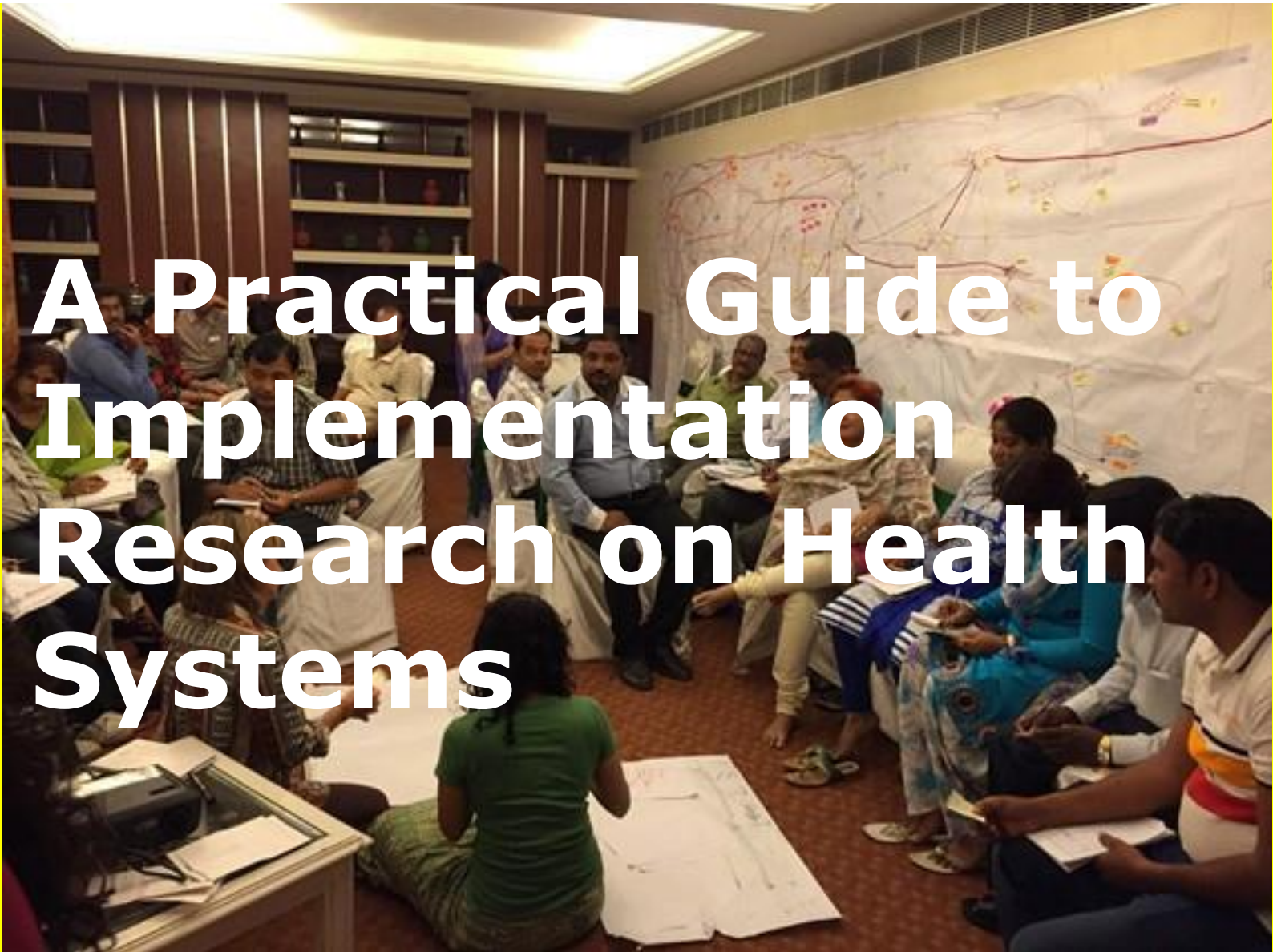# Future Health Systems
Innovations for equity

A Practical Guide to Implementation Research on Health Systems

Henry Lucas and Merrick Zwarenstein

IDS Institute of Development Studies

ARCADE HSSR
Building Capacity for Health Systems and Services Research in Africa

ARCADE RSDH
Building Capacity for Research on Social Determinants of Health in Asia

A Practical Guide to Implementation Research on Health Systems

Henry Lucas and Merrick Zwarenstein

Acknowledgments

Table of Contents

Chapter 1: Introduction

Henry Lucas
Institute of Development Studies

In his later writings the philosopher Wittgenstein concludes that in most cases 'the meaning of a word is its use in the language' (Wittgenstein 1958). Loosely interpreted, the advice is not to worry excessively about dictionary definitions or, in the present instance, about seeking to establish precise dividing lines between different research activities. We will try to follow that advice in this text by accepting that a very wide range of contexts and activities have been and should continue to be seen as within the boundaries of both 'Health Systems Research' (Gilson et al. 2012) and 'Implementation Research' (Peters et al. 2013a, Fixsen et al. 2005). We will set out our use of these terms in this chapter but make no special claims for the value of our interpretation over the many others that can be found in the recent literature.[i]

1.   Health systems

Following the above guideline, we would consider research on the impact of reforms to the UK National Health Service (Allen 2013), on the diverse range of formal and informal health providers in Bangladesh (Ahmed et al. 2013) the activities of patent medical vendors in Nigeria (Beyeler et al. 2015), or on community-based insurance in Laos (Alkenbrack and Lindelow 2013) as mainstream examples of health systems research. We would also include research on household heathcare-seeking behaviour (e.g. Diaz et al. 2013) or coping strategies in response to illness (e.g. Rahman et al. 2013). To what extent can we characterise these very diverse studies as research on 'Health Systems'?

The WHO defines a health system as:

> *"all organizations, people and actions whose primary intent is to promote restore or maintain health. … It includes, for example, a mother caring for a sick child at home; private providers; behaviour change programmes; vector-control campaigns; health insurance organizations; occupational health and safety legislation. It includes inter-sectoral action by health staff, for example, encouraging the ministry of education to promote female education, a well known determinant of better health."* (WHO 2007:2)

This is an interesting definition in a number of respects. First, it omits words such as 'interacting' or 'interdependent', which we would find in most dictionary definitions of the word 'system', though the same document later stresses the importance of interactions between health system components or 'building blocks'. Second, it defines the 'system boundary' - which organisations, people and actions are considered part of the system and which not - in terms of their 'primary intent'. This seems a potentially elusive criterion and one that must essentially involve a subjective judgement. Should we be examining, for example, the constitution of the Indian Ministry of Health and Family Welfare, the professional codes of conduct of Cambodian doctors or the mission statements of Ugandan Health NGOs to assess their health system status? Moreover, if individuals or organisations are contributing to the promotion, restoration or maintenance of health, to what extent does it matter why they are doing so? On the other hand, should we include those who have the best of intentions but who, perhaps because they have limited health knowledge, do more harm than good?

In line with the 'primary intent' requirement, health systems themselves are seen by the WHO as having goals, specifically, "*improving health and health equity, in ways that are responsive, financially fair, and make the best, or most efficient, use of available resources*" (WHO 2007:2). As might be expected from a United Nations agency, the overall approach seems most

appropriate for a 'national health system' with a defined organisational structure that operates under the direction of a well-intentioned government. Over recent years the WHO has expended considerable energy on devising methodologies to assess the comparative performance of such systems (Murray and Evans 2003) and producing guidance on 'health systems strengthening', which tends to focus on the central role that should be played by governments in terms of strategic planning, regulation and accountability, if not necessarily in service provision.

Though private providers are specifically identified in the list of organisations within the health system, it is not at all clear which, if any, private sector actors might be included within the WHO definition if strictly applied. Is the 'primary intent' of an international pharmaceutical company to 'promote, restore or maintain health' or to meet the expectations of its owners for a substantial return on their investment? In many countries there is a legal obligation on the directors of all companies with shareholders to act in the 'best interest' of those shareholders. It would seem unlikely that those interests would always align with the WHO criterion. Similarly, patent medical vendors in Nigeria are the main providers of anti-malarial drugs to the rural population, even though the law prohibits them from doing so (Goodman et al. 2007). They are typically marginalised workers on very low incomes whose 'primary intent' will in almost all cases be to support themselves and their families by responding to the demands of their clients. On the other hand, they would probably claim, and often genuinely believe, that they are proving a useful service to those clients, who have very limited access to any formal, regulated health services.

The current focus of WHO work on health systems is on the need for them to deliver 'Universal Health Coverage' (UHC):

> *"ensuring that all people have access to needed promotive, preventive, curative and rehabilitative health services, of sufficient quality to be effective, while also ensuring that people do not suffer financial hardship when paying for these services"* (WHO 2016)

Again, note that the focus is on health system *outcomes*. This is a perfectly rational approach, given that the 'primary intent' of the WHO is not to advance our knowledge of health systems but to improve the health of the world's population and that the primary mechanism available to them is to persuade national governments to play a lead role in improving the overall system that is currently delivering health outcomes in a given country. The 'building blocks' methodology of the WHO Health Systems Framework (WHO 2010) is an extension of this strategy. The aim is to identify a number of essential functions of a health system in order to consider the extent to which that system is meeting, or is capable of meeting, performance targets in terms of:

1. Health services
2. Health workforce
3. Health information
4. Medical products, vaccines and technologies
5. Health financing
6. Leadership and governance

For example, under the health services building block, the system should deliver 'effective, safe, quality personal and non-personal health interventions to those that need them, when and where needed, with minimum waste of resources' (WHO 2010:3). Using this approach, health systems strengthening can be defined in terms of (a) determining the extent to which any given component is failing to deliver its expected outcomes, (b) analysing the reasons for that failure - which may lie in its interactions with other components - and (c) implementing actions that will remedy the situation. Again, it seems evident that government will play the

key role in this process, possibly in collaboration with international agencies where resources are highly constrained or the national capacity for health systems strengthening is limited.

As indicated, we would see the above approach to health systems as one that may well be appropriate for the aims and procedures of the WHO. However, here we are concerned with health systems *research*. Gaining knowledge as to how specific health systems work is *our* primary intent. We assume that those working in this area will wish to use their research findings to influence policy in such a way as to 'promote, restore or maintain health', and a later chapter will provide guidance as to how this may best be achieved, but our first priority is to *understand* the health system that is the focus of our research - gathering and interpreting evidence about the complex interplay between the various actors who are engaged in what we identify as health-related activities (Peters 2014).

An interesting illustration of the possibilities for alternative approaches to the analysis of health systems is provided by Ahmed et al. (2013) in a study of the health sector in Bangladesh. The context within which that study is set will be very familiar to those who have worked on health systems in resource-poor environments. There are a multiplicity of health providers offering a variety of allopathic and alternative treatments, in this case including Ayurvedic, Unani and homeopathic remedies. Transactions are typically on a 'cash-for service' basis, even in the public sector. Poorer clients have very limited access to qualified providers (doctors, nurses, midwives, pharmacists) and rely on unlicensed village doctors, drug sellers, traditional healers, community health workers and traditional birth attendants. In this environment, the formal health sector regulatory framework has limited relevance for the great majority of the population and there is a yawning gap between the formal 'Bangladesh Health System', as defined in government policy statements, and the reality on the ground. To address this reality Ahmed et al. develop a conceptual framework that:

> *"challenges static and antiquated notions of policy and governance identified, for example, in the building block approach of the WHO Health Systems Framework or in the efforts to align development partners around a single country health plan. The complex and chaotic nature of health systems is unlikely to be tamed by these relatively naive notions of command and control health systems governance." (Ahmed et al, 2013:1753)*

Two relatively recent approaches to health systems have had a considerable influence on research in this area. The first is usually described as work on 'health markets'. At one level it involves an exploration of the role of private health providers, driven by a recognition of the extent to which in many countries health services are purchased in the same way as other services and commodities: "*in at least 19 countries in Asia and 15 countries in Africa − including many of the world's most populous nations (Bangladesh, China, India, Nigeria, and Pakistan) − more than half of total health expenditures are private out-of-pocket transactions*" (Lagomarsino et al. 2009:2). More generally, it questions the continuing relevance of many of the standard ways of characterising the health sector, recognising that health systems have become increasingly pluralistic. Old barriers between private/public, modern/traditional, and formal/informal health providers seem to be breaking down (Bloom et al. 2014; Peters and Bloom 2012). Bloom et al. (2008:2077) suggest that health systems can be more usefully considered as complex "*knowledge economies which produce and mediate access to health knowledge embedded in people, services and commodities*". Their work focuses attention on the 'stocks and flows' of health knowledge: how its value is determined, who possesses it and how others gain access to it. This requires a shift from traditional health systems analysis and its concerns with public and private sectors, modern and traditional providers, etc. and focuses attention on power relations and the ways in which it might be possible to construct new forms of "*social contracts for health care which build on existing areas of competence and good practice, whether mediated by states, markets or other institutional actors*" (Bloom et al. 2008: 2085).

The second approach, which has captured the imagination of many leading health system researchers over recent years, is based on the observation that health systems have all the characteristics of complex adaptive systems (CAS) (Carey el al. 2016, Bloom 2014; Rickles et al. 2007; World Bank 2007). A wide range of actors with diverse objectives act at multiple levels and interact through dynamic and multifaceted networks. As de Savigny and Adam (2009) point out, an intervention in one area will typically have consequences, often unforeseen, for many others.

> *"every health intervention, from the simplest to the most complex, has an effect on the overall system. Presumably simple interventions targeting one health system entry point have multiple and sometimes counterintuitive effects elsewhere in the system"* (de Savigny and Adam 2009:30).

Complex adaptive systems have the capability to self-organise, adapt, and learn from experience. They can change in a highly non-linear fashion over time, and are not easily controlled or predictable. It is not unusual for a CAS to show limited responses to apparently major interventions but then to change suddenly when a tipping point is reached (Gladwell 2000). In Chapter 4 we will discuss various aspects of CAS phenomena that are relevant to the analysis of health systems, including path dependency, feedback loops, scale-free networks, emergent behaviour, and phase transitions or tipping points (Paina and Peters 2011).

2.   A general framework for research on health systems

To provide a framework for the development of research strategies that can be used in exploring the wide range of health systems referred to above, we will adopt a generic approach to their definition, adapted from that proposed by Bishai (undated). We first define as 'agents' all those individuals who are considered to play a role in a given health system (doctors, nurses, managers, drug sellers, patients, carers, etc.). These agents may come together with a common purpose to form various identifiable 'units' (organisations or groups − ministries of health, hospitals, health centres, health insurance agencies, unions, households, etc.), which can be regarded as both capable of making decisions and responsible for any actions undertaken as a result of those decisions. For example, we might hold an individual hospital doctor responsible for their poor treatment of a specific patient but hold the hospital management responsible for their collective failure to employ sufficient doctors. Finally, we can consider the rules or 'institutions' that govern or at least influence the behaviour of these agents and units. The term 'institutions' is used to cover not only the relevant legal frameworks that regulate the health system but also any established procedures, protocols, guidelines, codes of conduct, accepted behavioural norms, etc., that agents and units are expected to observe.

Given the above, we can define a health system very generally as 'an interacting collection of agents, units, and institutions concerned with human health'. Note that this definition can encompass national systems such as the UK NHS, the patent medical vendors in a given state of Nigeria, community-based insurance schemes, and rural households in Bangladesh who are coping with the impact of healthcare costs. We make no *a priori* judgement as to the benevolence of the individuals or organisations involved, or as to the virtue of the institutions that influence their behaviour. Our initial aim will be to understand how a given system operates, though usually with the implicit intention of identifying potential ways to improve that operation in order to generate better health outcomes. The above implies that in any given context there may be multiple ways to define health systems. Health systems are essentially conceptual models of reality. "The concept of a 'health system' is a heuristic device for understanding a complex reality. Analysts draw different boundaries around the system depending on the questions they are trying to answer" (Bloom 2014:161). The important

question is not the extent to which they precisely mirror that reality – all economic and social models involve drastic simplification – but the extent to which they are useful in predicting and explaining observable outcomes. Thus, the first task of a health systems researcher will be to decide how they will identify the types of agents, units and institutions with which they will be concerned and how to specify the system boundary.

In quantitative studies, for example those using questionnaire surveys or analysing routine data, it will often be necessary to make such decisions at the start of the research process. To give a simple example, if the 'units' to be surveyed include private clinics, you will need to specify how private clinics are to be defined and sampled - often a far from simple task. Similarly, if the questionnaire survey is to gather information on the institutional context, for example rules on incentive payments, you will have to decide on the importance of informal payments and the extent to which you will attempt to explore the behavioural norms that govern such payments. As we will discuss in Chapter 7, one advantage of qualitative studies is that such definitions can be allowed to emerge and evolve during the research activity, though this advantage has to be balanced against the need to defend against challenges of subjectivity and bias that you will almost inevitably face from those who wish to dispute your findings.

Whether the health system under discussion is defined before or during the research activity, we would argue that the primary obligation on any researcher is for *transparency* - they must go out of their way to ensure that those who read, and may even use, their findings fully understand the assumptions made in arriving at those findings. Good researchers should have the confidence to expose themselves to critical evaluation of both their conceptual models and their methodologies, especially if they have expectations that their work may have a significant influence on the formulation of health policy.

3.   Implementation research

In his first address as incoming President of the World Bank Group, Jim Kim identified the 'next frontier' for the Group as:

> *"helping to advance a 'science of delivery'. Delivery isn't easy - it's not as simple as just saying 'this works, this doesn't'. Effective delivery demands context-specific knowledge. It requires constant adjustments, a willingness to take smart risks, and a relentless focus on the details of implementation".* (Worldbank 2012)

Implementation research (IR) can be seen as the means by which we can develop a 'science of delivery'. It has been defined as: "scientific inquiry into questions concerning implementation – the act of carrying an intention into effect" (Peters et al. 2013a). In the health sector, the focus on implementation research has arisen partly from a long-standing sense of frustration that interventions for which there appears to be strong evidence indicating the potential for substantial reductions in levels of morbidity and/or mortality in high-risk populations are either not being used or not being used effectively (e.g. Bhutta et al. 2014; Darmstadt et al. 2005). The primary objective of IR in health is therefore seen as the effective and efficient integration of such innovations into existing health systems, "to improve the uptake… of research findings into routine and common practices" (Padian et al. 2011:199).

As indicated above, there are multiple definitions of IR, often arising from the specific concerns of those working in different areas of health research (TDR 2016, WHO-COPR 2014). For example, those concerned with innovations in medical science, such as the development of new pharmaceuticals, will often use the term 'Translational Research' (Drolet and Lorenzi 2011) in relation to the overall process by which those innovations move from the laboratory to various stages of clinical trials on human subjects and then on to clinical practice. In the

present text, because we are focusing on IR in the context of research on health systems, we will be concerned with the final phase in this process, the "integration of research findings and evidence-based interventions into health care policy and practice" (NIH Fogarty International Center). We therefore exclude discussion of laboratory research to develop new drugs or medical technologies, and clinical trials to test the efficacy and safety of those drugs or technologies.

Our overall concern is thus to determine how best to apply health *innovations* that have proved successful in carefully controlled environments (laboratories, clinical trials, small pilot exercises, etc.) in a wider context. This requires the design of some form of *intervention*, which we will use as a general term to cover a range of activities including policy changes, programmes and projects. These have to be *implemented*, which typically involves actions by a collection of individuals that will here be described as the *implementation team*. In general we will assume that IR is best undertaken by 'insiders' – here defined as individuals who work *alongside* the implementation team, though with their own terms of reference and independently funded - and that the research questions they address are generated by identifying the constraints and challenges encountered during the implementation process. The scope of IR studies and the range of issues addressed can be very wide, including "the factors affecting implementation…, the processes of implementation themselves…, and the outcomes, or end-products of the implementation under study" (Peters et al. 2013b: 27).

Any implementation will take place within the context of an existing health system, which is in turn embedded in a broader physical, social, economic, institutional and, often overlooked, historical context (Grundy et al. 2014). A myriad factors may thus impact on the relative success or failure of that implementation, the great majority of which will be outside the control of the implementation team. The most appropriate strategy will often be 'constrained adaptation' - modification of the intervention design to allow for contextual factors but not to the extent that the primary aims of the intervention may be subverted.

The distinction between IR and the closely related activity designated 'Operations Research' (OR) is hard to pin down. For example, Zachariah et al. (2009) define OR as "the search for knowledge on interventions, strategies, or tools that can enhance the quality, effectiveness, or coverage of programmes in which the research is being done" (Zachariah et al. 2009:711), which bears a close similarity to the definition of IR provided above. In practice, definitions vary from agency to agency. An important framework document from the Global Fund (2008) on IR and OR makes no attempt to differentiate between them, other than to provide an annex listing a selection of these definitions. In this text we make a pragmatic distinction, drawing on the following definitions:

OR is:
the use of systematic research techniques for program decision-making to achieve a specific outcome. OR provides policymakers and managers with evidence that they can use to improve program operations.
(WHO 2003:3)

IR should:
a. identify common implementation problems… b. develop practical solutions to these problems… c. determine… the best way to introduce these new implementation strategies into the health system and facilitate their full scale implementation, evaluation and modification, as required.
(TDR 2005:4)

Note that the first definition focuses on the use of research to improve the operation of a given programme - to improve the implementation of the programme within which the research is undertaken. It is not unusual for such a programme to have an OR component, with terms of

reference that require those working on this component to focus on research that project managers may be able to use to enhance the implementation process (e.g. Heart 2016). The definition implies that OR is really only useful if it suggests actions that can be undertaken by the project management. In contrast, the second requires the IR team to use its independently funded resources to propose implementation strategies that can be *integrated into the health system*, by exploring not only issues that might hamper the current implementation but those that might be encountered in other contexts. In doing so, they may well be able to feed those strategies back into the implementation in which they are embedded, playing an OR role, but their primary task is to look outwards and explore the broader implications of their research in terms of seeking to maximise access to the benefits of the innovation. We therefore adopt the following working definition of IR:

> *Research that can provide evidence as to the advisability of scaling-up a given health systems intervention, and most effective means of doing so, on the basis of experience with one or more implementations of that intervention.*

To take a simple example, if an innovative incentive scheme for community health workers in a given region was being subverted by demands from higher-level health officials to be included in the scheme, an OR solution might be to ask the head of the regional government to negotiate with those officials, if that individual were strongly supportive of the intervention. From an IR perspective, this would raise a number of questions about the possibilities for both sustainability - how long will that individual remain as head of government - and scaling up. For example, how likely is it that similar issues will be encountered in other regions? Are there plausible alternative strategies that could be employed where the regional head was not willing to intervene or perhaps took the side of the officials? To what extent could the incentive scheme be adapted to gain acceptance among health officials while still delivering most of the anticipated gains in terms of health outcomes?

We would argue that the role of IR in terms of encouraging potentially wide-ranging reforms to the operation of health systems has implications for both the overall approach that researchers should assume and the research methodologies and methods that they should adopt. The failure of a time-limited project or programme in a given location will involve a waste of valuable resources and may delay or hinder the introduction of a potential valuable innovation. The failure of a significant health system reform could have far more serious consequences, both in terms of the size of the investment involved and the number of individuals affected. Researchers whose primary objective may be to encourage the widespread uptake of a health innovation and to influence implementation practice at scale have to ensure that their recommendations are backed by the most rigorous and persuasive research findings (Shaxson 2005).

In Chapter 2, we will consider the overall nature of the innovation process and relevant aspects of research design. We argue that innovations, even those which are seen as purely technical, provide a golden opportunity to improve the processes and outcomes of healthcare, not only by the effective implementation of that particular technology into the health system, but by improving the delivery system to take advantages of the opportunities afforded by that innovation. We would suggest that the primary requirement is for an initial very clear decision as to what the innovation under consideration for implementation is intended to achieve and what this may require in terms of alterations to delivery systems, which stakeholders will be affected by its implementation, and what factors will tend to facilitate of hinder its implementation. We are assuming that all those involved will be primarily concerned with influencing healthcare policy and/or practice and that reaching consensus among stakeholders on the appropriate approach to implementation is key to success. This collaboration should be a starting point both for implementation planning and for research design - what approach to implementation and scale up is most likely to succeed in the circumstances in which you are working and what types of research are most likely to generate

findings that will be accepted into the implementation process? We argue that while research quality and rigour are essential there is also a need to integrate research into the overall process of implementation, from initial discussion to the institutionalisation of the innovation into the health system, and to do this in collaboration with key stakeholders to the extent possible.

Given the above, we suggest that there are four broad areas that require special attention. The first, which we will address in Chapter 3, is the need to systematically review and evaluate the relevant existing literature, which will almost always be far more extensive than most researchers assume. The second requirement, discussed in Chapter 4, is to develop an in-depth understanding of the intervention that is to be implemented. This involves detailed knowledge of each step in the overall process that is intended to deliver the intended benefits, the assumptions that are required to hold for this process to function as planned and the indicators that will allow those managing the intervention to monitor if the implementation is on track. Because implementation is a dynamic process that will invariably diverge from the original plan, it will also be necessary to devise strategies that allow the researchers to be aware of any important modifications to that plan, focusing on the extent to which these are linked to the failure of the original assumptions or contextual factors that had not been fully appreciated when the plan was devised. As indicated above, one key issue will be the extent to which *adaptation* risks impacting on expected outcomes by threatening the *fidelity* of the implementation - has adaptation changed any essential features of the intervention, are we in practice implementing a reform that differs substantially from the one intended? As previously suggested, we would argue that only 'insiders' - researchers fully engaged with those undertaking the implementation - can hope to comprehend the process at this level of detail.

The third area is that of context, which we will address in Chapter 5. Here, as discussed above, we start from an analysis of the health systems context, with the health system and its boundaries carefully defined by the research team. We then need to explore the broader physical, social, economic, institutional and historical context within which the health system is located to identify those factors that may potentially influence implementation outcomes. We would see institutional analysis and stakeholder analysis as central to this process. Experience suggests that too often these are seen as peripheral activities, with rote procedures generating simplistic findings that play little part in the implementation process. Finally, in terms of overall approach, we argue the need for early and close engagement with key stakeholders, including those who may eventually play a central role in the integration of research findings into routine practice.

In Chapter 6 we will then address another often neglected aspect of health systems research, that addressing ethical issues. While ethical concerns have played a central role in the design of clinical research studies, there is often an implicit assumption that non-clinical health sector research is exempt from the strict observance of ethical standards. We will demonstrate why this assumption is unacceptable. The ambitious nature of IR also has implications for data collection, analysis and interpretation, as we will argue in Chapter 7. We have argued that it should be independently funded but recognise that this will probably mean that it have very limited resources with which to pursue its very ambitious objectives. A key skill with therefore be to use those resources most effectively, which involves careful allocation between a potential wide range of research activities. One guiding principle should be that of transparency. The objective of initiating large-scale reform of some aspect of the health system implies a need to influence a variety of key stakeholders, many of whom will have limited knowledge of data collection and analysis procedures. The overriding obligation of researchers is to the population who may be affected if their findings are put into practice.

Seeking to win over policymakers and other stakeholders by exercising analytical or presentational skills that mask underlying data limitations is not the way to meet that obligation. This is not to imply that sophisticated analytical techniques, for example econometric

modelling, should not be used in IR, simply to take note that we should be very wary of persuading policymakers of the likely benefits of a reform based purely on the findings of such an analysis, given that they are very unlikely to be aware of its possible limitations and underlying assumptions. At a deeper level, the potential importance of IR findings requires that researchers be sufficiently reflexive that they do not 'fool themselves' into believing that they have fully understood the nature of the implementation process by the routine application of qualitative and/or quantitative analytical procedures. Understanding typically requires that much more time is spent in careful consideration than in the manipulation of data. Three broad approaches to these issues, labelled by convention 'qualitative', 'quantitative' and 'participatory', though we recognise the inherent limitations of such categorisations, are explored in Chapters 8, 9 and 10. Recognising the central importance of gender issues, in general but very obviously in research on health systems, chapter 11 then sets out the need for a systematic approach to data collection, analysis and interpretation that incorporates a gender perspective.

Finally, chapter 12 considers perhaps the most frustrating aspect of policy-oriented research, that of trying to engage those who are in a position to use your research findings. As indicated above, there is a primary requirement to ensure the quality of those findings and present them in a manner that can be generally understood and correctly interpreted. However, while undertaking high quality research may ensure journal publication, it is rarely sufficient to influence policy-makers or even to gain their attention. Regrettable as it may seem, "*it is now well recognized that policy is determined as much by the decision-making context (and other influences) as by research evidence*". (Oliver et al. 2013:1). To have any chance of being effective the policy-orientated researcher has to endeavour to understand the policy process and the perceptions and motivations of the various actors who determine that process. This can be a challenging and time consuming process, once again emphasising the need to engage with stakeholders at the earliest possible stage of an implementation and to sustain that engagement over time.

Overall, we would argue that as researchers we have too often claimed to understand how complex health system interventions function on the basis of relatively flimsy evidence, typically involving short visits to the field, limited interaction with key stakeholders and one or two cross-sectional surveys of providers and/or intended beneficiaries that manage to be both complicated and simplistic. The argument in this text, as echoed in a recent book on the evaluation of complex interventions (Patton 2011), is that to reach a position from which we are willing to pass judgement on the advisability of scaling up or relocating a health system intervention that may have significant implications for the health and well-being of the population, we need to take implementation research much more seriously than in the past and reconsider the amount of time, effort and resources that we have previously been prepared to allocate to the task. We would suggest that, while senior international researchers may be able to play a valuable supporting role, the level of commitment required to lead such research will most likely be found among younger researchers from the region in which an implementation is undertaken (Hasnida et al. 2016).

References

Ahmed, S.M.; Evans, T.G.; Standing, H. and Mahmud, S. (2013). Bangladesh: Innovation for Universal Health Coverage 2: Harnessing Pluralism for Better Health in Bangladesh. *The Lancet* 382: 1746-55. http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(13)62147-9.pdf

Alkenbrack, S. and Lindelow, M. (2013). The Impact of Community-Based Health Insurance on Utilisation and Out-Of-Pocket Expenditures in Lao People's Democratic Republic. *Health Economics*, http://onlinelibrary.wiley.com/doi/10.1002/hec.3023/abstract (accessed 3 March 2015)

Allen, P. (2013) 'An Economic Analysis of the Limits of Market Based Reforms in the English NHS', *BMC Health Services Research*, 13 (Supplement 1): S1 www.biomedcentral.com/1472-6963/13/S1/S1 (accessed 28 January 2015)

Beyeler N, Liu J, Sieverding M (2015) A Systematic Review of the Role of Proprietary and Patent Medicine Vendors in Healthcare Provision in Nigeria. *PLoS ONE* 10(1):1-21. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117165

Bhutta, Z.A., Das, J.K., Bahl, R., Lawn, J.E., Salam R.A., Paul, V.K., Sankar, M.J., Blencowe, H., Rizvi, A., Chou, V.B. and Walker, N. (2014). Can Available Interventions End Preventable Deaths in Mothers, Newborn Babies, and Stillbirths, and at What Cost? *The Lancet* 384: 347-70. http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(14)60792-3.pdf

Bishai, D. (undated). *The Building Blocks of Health Systems*. Future Health Systems. www.futurehealthsystems.org/online-course/module-1-the-building-blocks-of-health-systems.html (accessed 20 July 2014)

Bloom, G. (2014). History, Complexity and Health Systems Research. *Social Science and Medicine* 117: 160-61. https://www.researchgate.net/publication/264248823_History_complexity_and_health_systems_research

Bloom, G. and Standing, H. (2001). *Pluralism and Marketisation in the Health Sector: Meeting Needs in Contexts of Social Change in Low- and Middle-Income Countries*. IDS Working Paper 136, Brighton, UK: Institute of Development Studies

Bloom, G., Standing, H. and Lloyd, R., (2008). Markets, Information Asymmetry and Health Care: Towards New Social Contracts. *Social Science and Medicine* 66(10): 2076−87. http://www.sciencedirect.com/science/article/pii/S0277953608000592

Bloom, G.; Wilkinson, A.; Standing, H. and Lucas, H. (2014). *Engaging with Health Markets in Low- and Middle-Income Countries*. IDS Working Paper (443), Brighton, UK: Institute of Development Studies, www.ids.ac.uk/publication/engaging-with-health-markets-in-low-and-middle-income-countries (accessed 28 January 2015)

Carey G, Malbon E, Carey N, et al (2015). Systems science and systems thinking for public health: a systematic review of the field. *BMJ Open* 5:e009002. http://bmjopen.bmj.com/content/5/12/e009002.full.pdf+html

Darmstadt, G.L., Bhutta, Z.A., Cousens, S., Adam, T., Walker, N. and de Bernis, L. (2005). Evidence-based, Cost-effective Interventions: How Many Newborn Babies Can We

Save? *The Lancet* 365.9463: 977–88.
http://www.who.int/maternal_child_adolescent/documents/pdfs/lancet_neonatal_survival_paper2.pdf

de Savigny, D. and Adam, T. (eds) (2009). *Systems Thinking for Health Systems Strengthening*, Geneva: World Health Organization. www.who.int/alliance-hpsr/resources/9789241563895/en/ (accessed 28 January 2015)

Diaz, T.; George, A.S.; Rao, S.R.; Bangura, P.S.; Baimba, J.B.; McMahon, S.A. and Kabano, A. 'Healthcare Seeking for Diarrhoea, Malaria and Pneumonia among Children in Four Poor Rural Districts in Sierra Leone in the Context of Free Health Care: Results of a Cross-Sectional Survey', *BMC Public Health* 13:157. www.biomedcentral.com/1471-2458/13/157 (accessed 28 January 2015)

Drolet, B.C. and Lorenzi, N.M. (2011). Translational Research: Understanding the Continuum from Bench to Bedside. *Translational Research* 157(1):1-5. https://www.viictr.org/viictr/assets/File/Drolet_Tranlational%20Research_2010(1).pdf

Fixsen, D.L.; Naoom, S.F.; Blase, K.A.; Friedman, R.M. and Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. National Implementation Research Network, FPG Child Development Institute, University of North Carolina, http://nirn.fpg.unc.edu/resources/implementation-research-synthesis-literature (accessed 28 January 2015)

Gilson, L. (ed.) (2012). *Health Policy and Systems Research: A Methodology Reader*. Geneva: Alliance HPSR/WHO, www.who.int/alliance-hpsr/resources/reader/en/ (accessed 28 January 2015)

Gladwell, M. (2000). *The Tipping Point: How Little Things Can Make a Big Difference*. London: Abacus. https://www.youtube.com/watch?v=jXZ7iNGPVnA

Global Fund (2008). *Framework for Operations and Implementation Research in Health and Disease Control Programs*. The Global Fund. http://whqlibdoc.who.int/publications/2008/9292241109_eng.pdf (accessed 28 January 2015)

Goffman, E. (1956). *The Presentation of Self in Everyday Life*. New York: Doubleday.

Goodman, C., Brieger, W., Unwin, A., Mills, A., Meek, S. and Greer, G. (2007). Medicine Sellers and Malaria Treatment in sub-Saharan Africa: What do They do and How Can their Practice be Improved? *American Journal of Tropical Medicine and Hygeine* 77(6) (Supplement): 203–18. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2657822/pdf/ukmss-4163.pdf

Grundy, J., Hoban, E., Allender, S. and Annear, P. (2014). The Inter-Section of Political History and Health Policy in Asia - The Historical Foundations for Health Policy Analysis. *Social Science and Medicine* 117:150–59. http://www.ncbi.nlm.nih.gov/pubmed/25066947

Hasnida, Amalaia, Robert A Borst, Anneke M Johnson, Nada R Rahmani, Sabine van Esland, Maarten O Kok (2016). Making health systems research work: time to shift funding to locally-led research in the South. *The Lancet Global Health*. http://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X(16)30331-X.pdf

Heart (2016). *How to motivate and support improved practices in Infant and Young Child Feeding (IYCF) in Northern Nigeria*. ORIE Research Summary. http://www.heart-resources.org/wp-content/uploads/2016/05/ORIE-RSummary-IYCF.pdf

Lagomarsino, G.; Nachuk, S. and Singh Kundra, S. (2009). *Public Stewardship of Private Providers in Mixed Health Systems: Synthesis Report from the Rockefeller Foundation - Sponsored Initiative on the Role of the Private Sector in Health Systems in Developing Countries*. www.rockefellerfoundation.org/uploads/files/f5563d85-c06b-4224-bbcd-b43d46854f83-public.pdf (accessed 28 January 2015)

Murray, C.J.L. and Evans, D.B. (eds) (2003) *Health Systems Performance Assessment: Debates, Methods and Empiricism*, Geneva: World Health Organization, www.who.int/health_financing/documents/cov-hspa/en/ (accessed 28 January 2015)

Oliver, Kathryn, Simon Innvar, Theo Lorenc, Jenny Woodman and James Thomas (2013). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research* 14(2)1-12. http://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-14-2

Padian, N.S.; Holmes, C.B.; McCoy, S.I.; Lyerla, R.; Bouey, P.D. and Goosby, E.P. (2011). Implementation Science for the US President's Emergency Plan for AIDS Relief (PEPFAR). *Journal of Acquired Immune Deficiency Syndromes* 56(3):199−203. http://www.pepfar.gov/documents/organization/157942.pdf

Paina, L. and Peters, D.H. (2011). Understanding Pathways for Scaling Up Health Services Through the Lens of Complex Adaptive Systems. *Health Policy and Planning* 27: 365-73. http://heapol.oxfordjournals.org/content/27/5/365.full

Patton, M.Q. (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York: Guilford Press. http://betterevaluation.org/resources/guides/developmental_evaluation/applying_complexity_concepts_patton

Peters, D.H. (2014) The Application of Systems Thinking in Health: Why Use Systems Thinking? *Health Research Policy and Systems* 12:51 http://idl-bnc.idrc.ca/dspace/bitstream/10625/53976/1/IDL-53976.pdf (accessed 28 June 2016)

Peters, D.H., Adam T., Alonge, O., Agyepong, I.A. and Tran, N. (2013a). Implementation Research: What It Is and How To Do It. *BMJ* 347: f6753. http://www.bmj.com/content/bmj/347/bmj.f6753.full.pdf

Peters, D.H.; Tran, N.T. and Adam T. (2013b). *Implementation Research in Health: A Practical Guide*. Alliance for Health Policy and Systems Research. World Health Organization. www.who.int/alliance-hpsr/alliancehpsr_irpguide.pdf (accessed 28 January 2015)

Peters, D. and Bloom, G. (2012). Bring Order to Unregulated Health Markets. *Nature* 487.7406: 163−65. https://www.researchgate.net/publication/229074487_Developing_World_Bring_Order_To_Unregulated_Health_Markets

Rahman, M.M.; Gilmour, S.; Saito, E.; Sultana, P.; and Shibuya K. (2013). Self-reported Illness and Household Strategies for Coping with Health-Care Payments in

Bangladesh. *Bulletin of the World Health Organization* 91: 449-58. www.scielosp.org/pdf/bwho/v91n6/13.pdf (accessed 28 January 2015)

Rickles, D.; Hawe, P. and Shiell, A. (2007). A Simple Guide to Chaos and Complexity. *Journal of Epidemiology and Community Health* 61(11): 933-37. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2465602/

Shaxson, Louise (2005). Is your evidence robust enough? Questions for policy makers and practitioners. Evidence and Policy 1(1):101-112. http://dx.doi.org/10.1332/1744264052703177

TDR (2016). *Implementation Research Toolkit*. Special Programme for Research and Training in Tropical Diseases (TDR). WHO.

TDR (2005) UNICEF/UNDP/WHO/World Bank Special Program for Research and Training in Tropical Diseases (TDR). *Implementation Research and Methods (IRM): Guidelines for Preparing a Grant Application to the TDR Steering Committee for Implementation Research*. http://geneplot.plasmodb.org/publications/tdr/grants/workplans/pdf/ir_guidelines.pdf (accessed 28 January 2015)

WHO (2010). *Monitoring the Building Blocks of Health Systems: A Handbook of Indicators and Their Measurement Strategies*. www.who.int/healthinfo/systems/WHO_MBHSS_2010_full_web.pdf?ua=1 (accessed 28 January 2015)

WHO (2007). *Everybody's Business: Strengthening Health Systems to Improve Health Outcomes. WHO's Framework for Action*. Geneva: World Health Organization, www.who.int/healthsystems/strategy/everybodys_business.pdf (accessed 28 January 2015)

WHO (2003). *Expanding Capacity for Operations Research in Reproductive Health: Summary Report of a Consultative Meeting*. WHO, Geneva, Switzerland, December 10-12, 2001. Geneva: World Health Organization, www.who.int/reproductivehealth/publications/general/RHR_02_18/en/ (accessed 28 January 2015)

WHO (2000) *World Health Report 2000*, Geneva: World Health Organization

WHO-COPR (2014). *Implementation Research Toolkit Video*. http://multimedia.who.int/mp4/WHO-COPR_Implementation_research_toolkit_26MAR2014.mp4

Wittgenstein, L. (1958) *Philosophical Investigations* (translated by G.E.M. Anscombe), second edition, Oxford: Blackwell

World Bank (2007). *Healthy Development: The World Bank Strategy for Health, Nutrition, and Population Results*. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/2007/01/8348853/healthy-development-world-bank-strategy-health-nutrition-population-results

Zachariah, R., Harries, A.D., Ishikawa, N. et al. (2009). Operational Research in Low-Income Countries: What, Why, and How? *The Lancet Infectious Diseases* 9:711-17. http://www.thelancet.com/pdfs/journals/laninf/PIIS1473-3099(09)70229-4.pdf

Chapter 2: Innovation, implementation and evaluation in healthcare

Merrick Zwarenstein
Centre for Studies in Family Medicine,
Schulich School of Medicine & Dentistry,
Western University, Ontario

1. Innovation in Health Systems

"*..(R)ather than asking how research evidence can be made more influential, academics should aim to understand what influences and constitutes policy*" ([Oliver et al. 2014:1](#)).

In this chapter we will consider the nature of change in health systems, and how small innovations in technology or service delivery complement large changes in policy or program design to improve outcomes of care. We provide guidance to health systems evaluators and researchers to help you deploy your scientific skills to design, implement, and scale up innovation in healthcare technology or service delivery, and assess their impact on health care and health outcomes. We hope you as evaluators and researchers will use some of these ideas to influence decision-making in health systems, in order to improve the health care experience of patients, communities and providers, health outcomes and health system efficiency.

The majority of health innovation ideas do not progress into viable products, services or changes in healthcare delivery and the failures occur at every stage. Few of those that are successfully developed and pilot tested are implemented effectively and even fewer scale to their full potential and are institutionalized into common practice. The best studied area of innovation is drug development, but the process of developing a new drug takes on average 14 years and 2 billion US $, and yet fewer than 5 percent of these reach scale and are sustained (NIH, 2014). The proportion of successes is simply not known for technologies, health service delivery or policy changes but given the relatively low investment in their early stage development, compared to the investment in drug development, there is reason to think that the success rate might be even lower.

Even if you have a good idea and a good innovation that is supported by empirical science that is simply not enough; the health system is complex and good innovations alone will not be effective in real world settings. Successful development, implementation and scale up of health innovations is a multi-stage process that requires appraisal at every stage and it is a team sport that requires active exchange and collaboration among all stakeholders at every stage. Successful uptake of innovations appears to depend on the interests of the critical stakeholders including the innovators, end users and the decision makers. Scale up is influenced by the broader context including the social and physical environment, the health system, and the regulatory, political and economic environment. Successful scale up is also strongly dependant on the maturity of the innovation at the point when it is offered to the health system decision makers for consideration. All too often, the maturity of an innovation is overoptimistically assessed, a usually by the innovators themselves, who are naturally confident in their idea. This assessment of the maturity would be more realistic if based upon acceptability of the innovation to other stakeholders, the evaluation results up to that stage and characteristics of the innovation itself including its disruptiveness.

We propose an approach to innovation which explicitly encourages this collaborative judgment of the maturity of an innovation and consciously undertakes specific work to ensure that it is ready for 'prime time' when it enters widespread use.

Our suggested approach to innovation and its spread into the health system consists of several stages: development, pilot testing, implementation, scaling up and institutionalization. The approach is accompanied at each stage by very careful evaluation, in order to identify potential problems that the innovation will face and facilitate remedies before large investments are made in a potentially flawed solution.

Is there 'a' health system?

We often speak of our countries' as each having a health system, bringing to mind a large, coherent, rationally designed and managed organization. In such a health system we might well assume that most change is implemented through major policy initiatives led by national, provincial and local government health departments. But healthcare systems are not tightly coordinated or well integrated machines. It would be more accurate to think of health systems as consisting of multiple separate and uncoordinated elements in a spontaneously and rapidly evolving eco-system, each element with a unique history and well established ways of doing things. These parts intersect, overlap, collaborate and compete against a background of changing patterns of disease, demography and care delivery. In this constant and often contradictory flux, widely varying responses to change in need, demand, social forces, pattern of illness are implemented, some as policy, but many more simply as ad-hoc decisions on delivery of care in reaction one or other currently high profile problem.

Governments are only one of the many groups trying to shape the health system in their own interests. Others include professional organisations, producers and sellers of drugs and technologies, non-profit governance organizations running hospitals or long term care homes, advocacy groups for specific patient and disease issues, and last, but not least, citizens and their families, as individuals, interest groups and communities. These end-users of health care may favour different approaches to care, with dramatically different priorities and proposals for structuring health systems, depending on whether they are young or old, urban or rural, recent migrants or long established, wealthy or poor (and even how poor) and depending on whether or not they are ill, and if so, with what conditions.

It is important for Health services researchers to understand that such complex systems are not easy to improve, and that well intended changes to one aspect of care may produce unintended consequences for another part of the health system. With complex patterns of needs, and complex structures for responding to these needs, how do health systems decision-makers decide what care to provide, to provide, to whom, and how? Whether to prioritize health services for children, or the elderly, on chronic or acute infectious illnesses, on equity, access or coverage of the population, on quality of care or continuity? Let alone the many other questions arising, such as whether primary care should be delivered by nurses, physicians, some other category of health worker entirely or inter-professional teams?

Many issues influence health systems decision-making, and scientific evidence is only one element. This evidence might include a randomized trial or systematic review on what intervention works best to deal with a particular health or health care problem, or new survey data on the rapidly rising prevalence of a particular health problem or of a problem with equity, cost, quality or access. It might be focus group data describing the perceptions of a particular group of users of care, or case studies of successful quality improvement initiatives. These kinds of evidence can influence decision making in different ways. Sometimes the evidence is used as a post hoc justification for a decision that has already been taken. This rhetorical use of evidence may ignore contradictory evidence. Evidence may also be used substantively, as a coherent and comprehensive overview of options and evidence leads directly to a decision that is supported by the prior evidence. While this substantive use of evidence may sometimes be very influential, especially when supported by prominent, positive media coverage, more often the evidence is one part of the impetus towards action, or towards choosing among options for action.

How do healthcare systems evolve?

Innovation and change in health systems can be at large or at small scale. Large scale policies have a profound influence on health care systems, determining their overall structure, funding, activities, eligible users and the health conditions they focus on. While these broad outlines determine the context in which care is provided, it may not necessarily determine the detailed daily operation, which are a result of the multitudes of small delivery processes chosen to implement each major policy. We propose that influencing these details of how care is provided, irrespective of the broader system context, is where our readers' skills and efforts may have the most impact.

Whereas national or provincial policymakers covering large jurisdictions need to have bold policies visible to those who elect them, lower levels of their organisations and smaller jurisdictions tend to focus on smaller, more local, operational choices rather than large scale policies. For health decision makers in such a position, it is important to continuously develop innovations that are carefully focussed on their priorities, so that they can spread and scale up the best of the interventions that improve the delivery of care, incrementally. As these small (and thus low risk) innovations accumulate, as successful innovations are evaluated, and distinguished from failures, which are dropped, the multiple small improvements in several aspect of care can accumulate to make a large impact on the overall outcomes of patients.

National government focus on making changes to law, like Obamacare, the Canada Health Act of 1967, or the South African Ministry of Health's commitment in 1995 to a National Health Insurance System. These result in policies which have enormous impact on how the health systems of those countries are structured, and thus on what their health systems can and cannot do. For example, Canada's Health Act focussed on physician centred acute care rather than chronic care, and did not require reimbursement of care provided by other professionals such as dentists and physiotherapists. The Act offered little funding support for long term or home care, and funded hospitals but not than community based services (aside from family physician care) or ambulatory pharmaceutical provision for the elderly. Even though these choices are federal, they have strongly shaped the structure and functioning of the provincial health systems. Thus, many major features of health care in all provinces are similar, even though there is complete constitutional autonomy of provincial health departments in relation to how care is delivered. This autonomy and strong central influence means that there has been little coordination or learning of lessons from each other as provinces have individually tried to adapt their structure to the changing needs of an aging population, and to control the costs of intensive, hospital based care within the requirements of the Canada Health Act.

In South Africa, with a similar national/provincial structure, and greater socio-economic and epidemiological challenges, including both HIV/AIDS and chronic disease, the commitment to a National Health Insurance System has focussed debate and senior decision-maker attention on how to fund care, with the consequence that strategic innovation in national policies and programs and public pronouncements has tended to focus on infrastructure and financing, rather than on detailed development, implementation and evaluation of delivery mechanisms. As in Canada, this has left the implementation of care delivery in the hands of provincial health departments, rather than the national government, with autonomy allowing locally relevant innovation. Although there is more communication and learning between provinces in South Africa than in Canada, there is similar influence of the national health policy priorities and structures on provincial priorities in care delivery, and similarly slow progress in designing large scale policies to deal with priority problems and demographic and disease challenges, such as chronic disease and HIV/AIDS at provincial level. In most countries, irrespective of level of income, strategic innovations in the form of high stakes national policy decisions with huge impact on the structure of healthcare occur infrequently, but occupy most of the attention of decision makers and the public,

Smaller scale innovation opportunities arise much more frequently. These arise most often where incremental changes in specific health care delivery mechanisms are need in response to a locally recognised problem, without major change in policy. Of course, innovation opportunities may also arise when detailed implementation plans for large-scale responses to growing health problems (e.g. chronic diseases, HIV/AIDS). The high level policy responses to these priority health problems rarely includes the detailed design of service delivery, leaving opportunities for health systems researchers to use creativity and scientific evidence in ways that are potentially less constrained by political requirements or rhetorical commitments than would be the case with lthe overarching policy. These seemingly "minor details" are fruitful work opportunities for Health Systems Researchers. Small innovations and improvements in healthcare may also arise in response to newly available (or, in low and middle income countries, newly affordable) technical innovations in prevention, diagnosis and treatment. When these technical innovations and combined with carefully designed changes in organization and delivery of care, they can, if they are well evaluated, successfully implemented and scaled up, become the basis for improved health systems, whether in high, low or middle income settings.

An incremental approach to health system improvement might be especially appropriate for constrained economies in this economically depressed stage of globalisation, where economic concentration and weakened social solidarity leads to shrinking states and public budgets. In spite of this apparent association with economic recession and spending constraints, this 'low' road to healthcare improvement can tap into the new knowledge generated by health service and systems researchers, in potentially advantageous ways. Technical innovations in prevention, diagnosis and treatment, in conjunction with finely tuned changes in organization and delivery of care, can, if well designed, be the basis for improved health systems, whether in high, low or middle income settings.

When deeply integrated into existing health systems, incremental innovations can result in unexpected positive consequences. When we designed a new training system for nurses in primary care clinics in South Africa to improve their ability to diagnose Tuberculosis (TB), we hoped only to improve the reliability with which tuberculosis would be diagnosed and referred for treatment. We had also a vague hope that this would demonstrate their capacities as clinicians, and open up a larger and more effective role for such nurses in publicly funded primary care in South Africa. Fifteen years later, over 20 thousand nurses are making use of a wide range of newly acquired skills to diagnose and treat not only Tuberculosis but a full range of minor acute and major chronic illnesses including hypertension, asthma and AIDS, using evidence based guidelines, all arising from the expansion of an incremental improvement- the development of effective and efficient in- service, on-site training systems. This success suggests that as primary care becomes an ever more important part of the health system, iterative improvements in organising care, delegating functions, sharing care and referring patients for specialized treatment can lead to improved coverage, quality and impact of care. This success is an argument for health service research to focus on implementation of incremental improvements to existing programs of primary care, using strategies that will be easily scaled up in existing local health systems without disruption.

The combination of incremental innovation with health services research, especially implementation research can help existing health systems evolve to deal with changing health and demographic trends while improving health outcomes, promoting equity and containing expenditure increases. Incremental innovation promotes simplification of care, and thus improves access to effective treatments or preventive interventions. In large part, the fall in child mortality throughout the developing countries since the 1960's has been due to the delivery by alternative, non-physician providers of simple and highly effective treatment or preventive interventions such as immunizations. It is easy to forget that smallpox was a world scourge, eliminated by a simple new vaccination technology, the bifurcated needle, and

delivered through an equally simple but well organized effort to isolate cases and immunize protective perimeters of populations around them. Similar effects have arisen from the development of effective treatments for chronic diseases including hypertension and diabetes, tuberculosis and HIV/AIDS, where innovations in treatment have been delivered alongside refinements and simplifications of care systems so that access, adherence and quality of care improve, and combine to reduce morbidity and mortality form these conditions.

2. The role of researchers in incremental innovation

Innovation is a long and complex process and thus highly unpredictable; health systems researchers can help ensure that as an innovation is developed, implemented, spread and scaled up to cover entire jurisdictions, that it is evaluated carefully at each stage, and that the lessons learned from that evaluation inform either early abandonment if it is clearly not an effective innovation, or improvement, to ensure that a successful innovation can be most easily implemented and can achieve its maximum impact. This continuous attention to evaluation and iterative improvement (or appropriate abandonment) reduces the unpredictability of innovation.

Health systems research offers many insights during the complicated process of innovation: HSR helps understand the problem that needs to be solved, ensures that the proposed innovations are acceptable to those who will be affected by their implementation, that the chosen innovation achieves its expected benefits, and that it does not create any unexpected harms or costs, either within the part of the system in which it is implemented, nor in other parts through unexpected links. Researchers provide information and evidence which can assist decision makers in several ways:

- defining the priority problem to be addressed;
- designing and choosing among options for the innovation;
- developing and testing implementation and scale up strategies;
- determining the impact of real world implementation on health and healthcare delivery; and
- recognising areas for improvement in future iterations.

We will discuss methods and timing of evaluation in incremental innovation. Evaluation is crucial, because it ensures that each of these innovations is indeed an improvement, and not simply added work and cost with no benefit, or, even worse, no benefit and extra harms and/or costs. If an evaluation shows that the innovation is not effective, then it is possible to stop the innovation, not implement or scale up widely, and rethink, with modifications to the innovation based on flaws recognized in the evaluation process. This iterative cycle of innovation, implementation, evaluation, improvement, and so on to another round of innovation, is a key approach to making a difference with health systems and services research

With the complexity of the health system described above, it is best for health systems researchers to collaborate closely with other stakeholders to succeed in tactical healthcare improvement, either in an actual team (within an organisation like a ministry of health or a non-governmental organisation) or in a virtual team (if different stakeholders from different organizations are working together on an initiative). This collaboration goes through several steps as discussed below.

3. The stages of innovation: from problem to scale up

The most commonly cited ethical principle that healthcare students are taught in medical schools is 'First, do no harm', which for our purposes might be interpreted as 'be very aware

of the potential risks associated with any health systems innovation as well as the potential benefits'. The way in which we can best try to ensure that is to proceed with caution, moving step-by-step through the following stages:

1. Problem: identify priority problems that are susceptible to tactical solutions
2. Solution: develop one or a few potential solutions to the point where they can be tested in a small, real world pilot or identify plausible solutions developed elsewhere and adapt them to local conditions
3. Pilot test: if it tests well then prepare for larger implementation, if it needs improvement, adapt and pilot test again, if it seems not to be improvable, abandon
4. Implement: in similar settings to pilot, but at larger scale and under real world conditions. Evaluate and decide on spread to different problems and settings, and/or on scale up to jurisdiction/s
5. Scale up: based on evaluation of implementation stage, adapt the innovation and supporting systems to allow massive growth, and test whether it can be adapted to solve different problems, or the same problem in different settings. Evaluate jurisdiction wide scale up, especially whether effectiveness has been maintained, with rigorous, often randomized longer term evaluations of results and implications for other parts of health system.

An innovation's ability to progress through these stages is contingent on several factors (Gupta, 2015). It is dependent on the characteristics of the innovation itself and the interests of the key stakeholders including:

- innovators (usually researchers) who are involved in developing the innovation;
- end users (the practice community and innovation users) from the health system unit;
- decision makers (government and non-government policy makers) who have policy jurisdiction within the health system unit.

It is also dependent on the broader context including:

- the social and physical environment,
- the health system unit where the innovation will be integrated (i.e. organization, clinic, hospital, community, province etc.); and
- the regulatory, political and economic environment.

It is important to identify barriers early in the innovation process and accept that some innovations simply may not be able to overcome important barriers and perhaps, there will be a need to go back to earlier stages and re-design the innovation, or in some cases abandon the project all together. Innovations are commonly rushed through stages and even skip essential stages all together. They may be implemented or scaled up prematurely without evaluations to verify that they are mature enough to advance forward.

Open and thoughtful (rather than rhetorical) discussion is needed between multiple stakeholders, including health innovators, decision makers and end users on potential barriers to scale up as they come into view, allowing for innovations to be sequentially adapted before meeting these problems in the "real world" setting. Collective problem solving among stakeholders is an essential element of deliberation, which "*allows individuals with different backgrounds, interests and values to listen, understand, potentially persuade and ultimately come to more reasoned, informed and public-spirited decisions*" (Abelson 2003:241).

It is helpful to be constantly aware of what *stage* the innovation is at and to identify what barriers have to be overcome in order to move forward in the process of implementation and

scale up. Awareness from the beginning of the whole process leading to the end stages increases the ability to pre-empt barriers, and the likelihood of achieving successful scale up and spread of an innovation. This approach to staging of innovation may be most usefully applied to discrete innovations and to multicomponent interventions, rather than paradigmatic innovations ([Edwards 2010](#)). Paradigmatic innovations are often attempted as solutions to difficult strategic problems, and as discussed above, these may be easier to solve in a piece by piece fashion.

Discrete innovations are well defined such as scale up of zinc in early childhood ([Larson et al. 2012](#)), scale up of ART ([Harries et al. 2009](#)) or the use of new technology for diagnosis and treatment of TB ([Meyer-Rath et al. 2012](#)). Multicomponent interventions involve several interacting program elements to produce a composite set of innovations that are targeted at multiple system levels. Examples include multilevel initiatives to decrease childhood obesity ([deSilva-Sanigorski et al. 2010](#)) or scale up of post abortion care services in two countries (Billings et al. 2007). Paradigmatic innovations require a shift in the way we understand health problems and the potential solutions to address them. An example of this is China's quality of care reforms to modify their family planning programs to be in line with the international agenda which required a systems wide approach, and partnerships between international groups and all levels of governments in China, including those that extend outside of public health ([Kaufman et al. 2006](#)).

Stage 1: Identify the problem to be solved
If a problem is widely discussed, its characteristics understood and magnitude well measured, its priority agreed upon by the full range of stakeholders, including those with the health problem, their communities, the professions and organizations providing care and health care funders and decision makers it is likely that health systems research skills can help to address it. It is easier to tackle if the problem has a high public profile and a solution is required by new laws (or at least, not prevented by any), or is enabled by a newly available technology or healthcare delivery change.

Difficult problems have more multifactorial origins, are deeply rooted in cultural, social or economic stresses, have more polarization, stigmatization, or conflicting interest groups. Perhaps a chain of simple innovations can help, building up over time, with each small step addressing one small part, and achieving gradually widening support. Often some of these stages require new laws, new financial commitments and complicated political support. These difficult, strategic problems are often the most important problems in health systems, the result of inequitable social situations, but taken as a whole, such problems are hard to solve in one step. We suggest that you try to work on a mixture of simpler and more complex problems, preferably related to each other so that the learning you achieve from one helps you to understand and possibly help with others.

Stage 2: Find or develop a solution
New innovation development should only progress if it is clear that there are no existing solutions to the problem. Do a literature search to check whether this problem has been addressed elsewhere, and, if so, how. Ask your networks if they know of existing solutions. If not, start thinking about the innovations needed to solve this problem. Use an approach called user centred design ([LeRouge & Wickramasinghe 2013](#)). Consider if the innovation should be aimed at clinicians, managers, a team, multiple units or facilities, jurisdictions or end users, and which ones? Does it focus on individual awareness, knowledge, motivation, attitudes, engagement, skills, behaviour or work processes? Is it a drug, a technology or a process change? Diagram how you think an innovation will work. Gather feedback from end users and managers to test your assumptions about what is needed and to direct the design. Develop one or a few local innovations- keep them simple, adaptable by end users and compatible with the existing culture, health system and workflow. Rapidly test alternative ideas with enthusiastic users, starting with simple pictures of the solution, moving to physical mock-ups

and/or role-plays of the innovation in use. Go through several cycles of prototyping, feedback, adaptation until ready for pilot testing.

Stage 3: Pilot test

Test a real version of the innovation with a few local enthusiasts, who are ordinary end users (patients, communities, providers) in a real setting. Evaluate convincingly, using transparent qualitative and quantitative measures: is it acceptable to all stakeholders, does it work, is it simple, does it integrate into the system easily, is it better than the alternatives, at what cost? If not, abandon, or improve the innovation. After improvements, test again. When there are no major uncertainties, get ready for implementation.

Stage 4: Implement

In similar settings to pilot, but at larger scale, under real world conditions and with comparative effectiveness evaluation built in. Consider contracting out for implementation and consider recruiting independent evaluation team. In any case, build an implementation and evaluation team with buy in from end users, including patients, providers and communities, local respected champions, decision makers from several levels and strong administrative support; also advisers with KT knowledge. Ensure shared implementation decision making between stakeholders, communications of progress and an agreed performance measurement framework based on logic model from previous stages.

Evaluation should be pragmatic, realist and participatory. Effectiveness should be measured both in processes (how has healthcare delivery changed) and in outcomes (how has health or other end user relevant outcomes changed). Designs should include rigorous, preferably randomized trials, with mixed methods (including trial, qualitative and economic) evaluation including satisfaction, user experience, uptake, quality, effectiveness, and economic measures and observations. Look for unintended consequences and system impacts especially opportunity costs of implementation and e.g. internal diversion of resources and performance decline in other areas of function of involved delivery organizations. Report on social, cultural, geographic and health system effects on innovation; consider regulatory legal and financial barriers and potential solutions. Report on external validity/generalizability as well as effectiveness, benefits and harms in different subgroups, and recommend whether the innovation is ready for spread to different settings, problems user groups, (spread) and/or whether it is ready for scale up (expansion of innovation to other but similar settings dealing with the same or similar problem. If not, make explicit whether adaptation is possible, and if so, along what lines, or if a new direction is preferred.

Stage 5: Scale up

Assuming the evaluation from the implementation stage is positive and recommends scale up adapt the innovation as suggested and choose new problems or settings, if extensions is to be addressed first; or identify an expansion path (similar settings, same problem, minimal adaptation) if the decision is that the innovation is able to scale, but not extend to different settings or problems. Adaptation is based on rethinking the logic model, to see which elements can and need to be changed to match the different situation or problem. Consider the core and adaptable elements, and how to adapt the latter for the different settings or problems while maintaining sufficient fidelity to the original successfully implemented innovation to continue to be effective. Scale up may require changes in the physical, health system or legal/regulatory/financial context in which the intervention is to be implemented,- possible changes include to the delivery mechanisms, capacity development, funding any of which may need to be further developed to assist in scaling up a successful innovation to similar settings on a jurisdiction-wide or multi-jurisdiction scale,

The stage of scale up needs to be evaluated as well, as thoroughly and rigorously as the implementation stage itself was evaluated. This is because, inevitably, the initial

implementation, like the pilot much earlier, is led by the most committed to the innovation, implemented in the site most likely to succeed, and reviewed through the most optimistic lens, by decision makers whose reputation is built on announcements of successful pilot projects being widely implemented. It becomes all the more important that the long term commitments on a massive scale that accompany a decision to scale up, with or without spread beyond the area and problem initially targeted, are based on a rigorous objective and possibly independent evaluation of whether or not the expected gains are actually forthcoming. A reliable evaluation of the initial efforts at spread and scale-up provides the ability to correct course in order to maximise the positive and minimise unexpected negative consequences, and the reassurance to all stakeholders, prior to setting the innovation into the system, irreversibly for the foreseeable future that the innovation deserves to be scaled-up and spread. This evaluation at scale must also consider implications of the scale up efforts on other innovations or other parts of the health system.

4. Conclusions

Health systems are complicated and improving them in ways that achieve wide and positive impact depends on understanding the particular problem which you want to solve very carefully. This may mean breaking down bigger problems into manageable pieces and developing innovations for each one, rather than trying to solve deep problems all at once. A new idea will not necessarily work, and even if it does do so at a small scale, innovation is not self-implementing. Each innovation needs to be tested and only if it is successful should it pass on to the next stage. A large part of successful innovation is knowing when something has failed, and not trying to scale it. If an innovation appears to be successful as a prototype in pilot studies, it should be tested in a larger scale, using rigorous evaluation tools; with this information, if positive, it is worth trying to adapt the innovation to try and spread it as a solution to other problems, or the same problem in other settings; and also to scale it up widely across jurisdictions. Even at this stage it remains important to evaluate, to see if the earlier successes are maintained at scale.

References

Abelson, J., Forest, P.-G., Eyles, J., Smith, P., Martin, E. and Gauvin, F.-P. (2003). Deliberations About Deliberative Methods: Issues in the Design and Evaluation of Public Consultation Processes, Social Science and Medicine 57:239-251. http://www.ibrarian.net/navon/paper/Deliberations_About_Deliberative_Methods__Issues_.pdf?paperid=3703492

Arendt, H. (1958). The human condition. Chicago: University of Chicago Press.

Billings DL, Crane BB, Benson J, Solo J, Fetters T. (2007). Scaling-up a public health innovation: a comparative study of post-abortion care in Bolivia and Mexico. Social Science and Medicine 64(11):2210-2222. https://www.ipas.org/~/media/Files/Ipas%20Publications/BillingsSSM2007.ashx

Edwards, Nancy (2010). Scaling-up Health Innovations and Interventions in Public Health: A Brief Review of the Current State-of-the-Science. Paper commissioned by the conference chairs for delegates of the inaugural Conference to Advance the State of the Science and Practice on Scale-up and Spread of Effective Health Programs, Washington, DC, July 6-8. www.ihi.org/education/Documents/ProgramMaterials/ScaleUpBlog/7a_Commissioned_Paper%202_Public_Health.doc

Gupta A, Thorpe C, Bhattacharyya O and Zwarenstein M. Promoting development and uptake of health innovations: The Nose to Tail. F1000Research 2016, 5:361 (doi: 10.12688/f1000research.8145.1). http://f1000research.com/articles/5-361/v1

Harries AD, Zachariah R, Jahn A, Schouten EJ, Kamoto K. (2009). Scaling up antiretroviral therapy in Malawi - implications for managing other chronic diseases in resource-limited countries. Journal of Acquired Immune Deficiency Syndrome 52 Supplement 1:S14-16. http://fieldresearch.msf.org/msf/handle/10144/88073

Kaufman J, Erli Z, Zhenming X. (2006). Quality of care in China: scaling up a pilot project into a national reform program. Studies in Family Planning 37(1):17-28. http://paa2005.princeton.edu/papers/50453

Larson CP, Koehlmoos TP, Sack DA (2012). Scaling Up of Zinc for Young Children (SUZY) Project Team. Scaling up zinc treatment of childhood diarrhoea in Bangladesh: theoretical and practical considerations guiding the SUZY Project. Health Policy and Planning 27(2):102-114. http://heapol.oxfordjournals.org/content/27/2/102.long

LeRouge, Cynthia and Nilmini Wickramasinghe (2013). A Review of User-Centered Design for Diabetes-Related Consumer Health Informatics Technologies. *Journal of Diabetes Science and Technology* 7(4):1039-1056. http://dst.sagepub.com/content/7/4/1039.full.pdf+html

Meyer-Rath G, Schnippel K, Long L, MacLeod W, Sanne I, et al. (2012) The Impact and Cost of Scaling up GeneXpert MTB/RIF in South Africa. PLoS ONE 7(5):1-11. www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0036966&representation=PDF

National Institute of Health (NIH). Clinical and Translational Science. Available at: http://www.ncats.nih.gov/research/cts/cts.html. Accessed August 13, 2014.

Oliver, Kathryn, Lorenc, Theo, Innvaer, Simon (2014). New directions in evidence-based policy research: a critical analysis of the literature. Health Research Policy and Systems 12(34):1-11. http://www.health-policy-systems.com/content/12/1/34

de Silva-Sanigorski AM, Bolton K, Haby M, Kremer P, Gibbs L, Waters E, et al. (2010). Scaling up community-based obesity prevention in Australia: background and evaluation design of the Health Promoting Communities: Being Active Eating Well initiative. BMC Public Health 12(10):1-7. http://www.biomedcentral.com/content/pdf/1471-2458-10-65.pdf

A useful resource: The Nose-to-Tail Tool

The Nose-to-Tail tool is intended to help stakeholders identify the stage of maturity of an innovation, facilitate deliberative discussions on the key considerations for each major stakeholder group and the major contextual barriers that the innovation faces. It should help to identify potential problems and facilitate early modification, before large investments are made in a potentially flawed solution.

Chapter 3: What do you know?

Introduction

In chapter one it was argued that implementation research should be an ongoing activity, tracking the progress of what will typically be a complex health systems intervention and attempting to build an understanding of both what is happening as the implementation evolves and, an even more difficult task, why it is happening. Interpretation of data clearly involves a reasonable degree of intelligence and an ability to think rationally about the interplay between intervention activities and the context within which they are played out. However, experience can be an equally important guide, both your own and that of the multitude of researchers and others who have gone through similar processes before you. Being able to identify, assess, assimilate and use relevant existing evidence that may provide valuable insights is one of the key attributes of a capable implementation researcher. In this chapter the focus is on the first two activities - locating relevant evidence and assessing its quality.

We can distinguish between two phases of evidence review. Initially, we will need to draw on the existing literature in the design of our research. It will help us both to refine our research questions and to develop the appropriate methodologies for data collection. A selective review of the recent literature will also be essential if, as advocated in chapter one, we seek independent funding for our research. Those offering funding will be expecting us to provide findings which will complement the existing body of knowledge on a given topic. They will need to be convinced that we are very familiar with that knowledge and that our research is targeting areas where evidence is currently lacking. The first part of this chapter describes the basic review process from this perspective.

In the second part we consider what can be seen as a natural extension of this initial phase, the undertaking of a 'systematic review'. This term is usually dated back to a book by Cochrane (1972), which argued that with limited resources available in the health sector, clinical judgements should be based on all the available evidence on treatments that had been obtained from rigorously designed evaluations. While that book, and the continuing work of the Cochrane Collaboration in this area, strongly emphasised the importance of one particular approach to evaluation - the Randomised Control Trial (RCT) (J-PAL undated) - many authors have suggested that, particularly when considering innovations not directly concerned with clinical trials, the range of material considered should be substantially expanded, while retaining two key features of the methodology: the aim of systematically compiling *all* the relevant literature; and the rigorous quality assessment of each item before incorporating its findings to the extent warranted by that assessment into a final overall synthesis.

Our suggestion here is not that every implementation researcher should conduct a systematic review, though a recent proposal goes further, arguing that, given the relative ease with which they can now be undertaken using the internet, there should be "*no new studies without adequate systematic review of existing evidence showing new research is justified*" (Lund et al. 2016:5). The article points to at least one major research funding body which has accepted this policy. Our more modest suggestion is that if researchers are going to have a long term involvement with an implementation of a given intervention, it would be advantageous to allocate some of their time to following a process similar to that required for a formal systematic review. By defining appropriate selection and assessment criteria for such a review, given the nature of the intervention with which they are engaged, it may be possible to refine their interpretation of the data they are compiling by building systematically on the experience of researchers who have addressed similar issues.

Part 1: Rapid literature reviews

Catherine Grant
Institute of Development Studies

1.   What is a literature review?

A literature review should include a select analysis of existing research that is relevant to what you have been asked for in the application, showing how it relates to your proposed research. It explains and justifies how your investigation may help answer some of the questions or gaps in this area of research and promote your application as a necessary area of study. A literature review is not a summary of everything available on a specific topic and it is not a chronological description of what has been discovered about a particular area. It is important to be concise, clear and selective, especially when writing a review for a funding application, bearing in mind that the people reading the application may not be experts in this issue, so avoiding any acronyms or very specific language.

If you are seeking funding, first check the donor criteria for their support and show how your project fits. Such is the competition for funds that there is no point in submitting a project, however worthy, if it does not clearly meet donor priorities. There are different types of funding applications and the amount of evidence you will need for your literature search will depend on what they are asking for so clearly read this before going any further with your search. One common way to approach the structure of a literature review is to start out by outlining the context and then become more specific, as suggested in figure 1 (University of Reading undated). First, explain the broad issues related to your research proposal; this should not be too long, just enough to explain the context. Next, focus on studies in your particular area of research, followed by those directly relevant to that research and particularly those that identify gaps in the literature.

Figure 1

Broad Issues

Overlapping Studies

Directly Relevant Studies
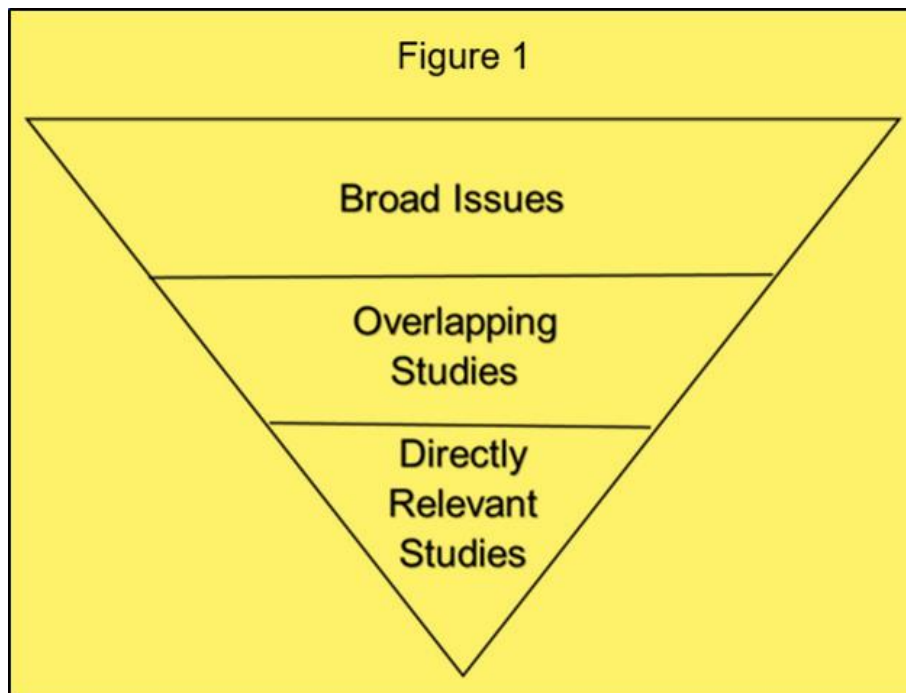
2.   Search strategies

Identify a research question

Start with a carefully thought out research question which matches what the funder is asking for. A literature search should be focused and to ensure you are efficient with your search you must be clear from the start what types of evidence will be relevant to address that question. There are many guides that can help with this (e.g. Aveyard 2007, Chapter 3). A systematic approach to searching for the literature is key. Ensuring that you follow a structure will allow you to identify the key broad texts and find the specific studies that are most relevant to your work. It may help to break the literature search into key themes with different sets of keywords, to help with organising your search as suggested in the diagram above. Make sure you record how you have approached your search and if you have been short on time and had to adapt some of these processes for speed that is fine too.

Keywords

The keywords you choose are central to shaping your search. You will know some of the appropriate words but may need to use a snowball approach and add keywords as you access the literature and increase your knowledge of the terminology being used. If you are new to the topic do an extremely brief general search to help identify your keywords. You should be as creative as possible at this stage, as this will form the basis of your search and restrict what you find. You will need to consider that there are different meanings to different words, and also consider that different spellings and different terminologies may also be used in different countries. Note that keywords need not only relate to terms in your research questions. If your searches identify authors or agencies who have regularly published in the area, you can also search using their names.

> Example: Attitudes to medical abortion in India:
> - Overall search (broad, context-setting)
>   *Keywords: abortion, India, attitudes*
> - Theme 1: Medical abortion in the South Asian context (relevant studies)
>   *Keywords: medical abortion, Asia*
> - Theme 2: Personal characteristics affecting attitudes to abortion (relevant studies)
>   *Keywords: education, socioeconomic, parity, abortion, personal characteristics
>   (then add words in a snowball approach as you read through studies and find out what works)*
> - Theme 3: (specific): Attitudes to abortion in India
>   *Keywords: Identify keywords based on the information you have found from the other searches about what terminology is used.*

Take some time to get to know the search engines and how they work; for example exploring the use of AND/OR/NOT and * commands can be very useful when conducting your search and can save you time:

- AND ensures you search for two or more specified terms;
- OR looks for any one of them;
- NOT excludes articles with specific terms; and
- * Allows any ending to be searched for, e.g. anthropo* will bring up anthropological, anthropology, anthropologist, etc.

For example, table 1, shows various ways of refining a search on the links between hand washing by staff and hospital acquired infections.

Table 1: A search on links between hand washing and hospital acquired infections

Operator	Search	Retrieves
AND	hospital acquired infection AND hand washing	Retrieves citations with BOTH terms present
OR	hospital acquired infection OR cross infection OR nosocomial infection	Retrieves citations with ANY of these terms
AND, OR	(cross infection OR nosocomial infection OR hospital acquired infection) AND hand washing	Search sets may be combined. This search locates citations with the word hand washing AND (ANY of the terms combined with OR)
NOT	hand washing NOT masks	Retrieves citations with the term hand washing, but omits records with the term masks (Caution: the NOT operator should be used carefully as it may omit citations relevant to a search. For example, an article about hand washing that includes the word masks might be relevant.

Source: NYU Libraries (2016)

Identify types of literature to include:

Next, decide which types of literature you will include. This will help you to narrow down your search and also decide where you might best search for information. For example, you may want to include newspaper reports if you are looking at public opinion, or definitely exclude them if you are looking for an academic evidence base. Examples of the types of literature to include for the health sector are:

1. Peer-reviewed and academic journals using relevant search engines, e.g. Google Scholar, Scienceopen, PubMed (which provides free access to the MEDLINE database), and the WHO Library & Information Networks for Knowledge Database (WHOLIS).
2. Full text versions of journal articles available for selected countries using HINARI.
3. Working papers published by established research and consultancy agencies.
4. International and national policy documents.
5. Websites of international organisations, private companies and NGOs, and grey literature (NIH 2016) - newspapers, magazines, blogs, etc., often identified using Google or other general purpose search engine.

You need to spend time thinking about the advantages and disadvantages of using different sources. Academic articles and books should have been peer-reviewed, which provides at least some guarantee of quality. However, there are often considerable delays between preparation and publication, so they may not provide the most recent data. Reports produced by an international agency may reflect the specific objectives of that agency or be influenced by political considerations - for example not wishing to provoke a country that is contributing to its budget. This may be an even more important consideration for material produced by private companies and NGOs. Grey literature typically will not have been through a process of peer review and may well be seen by some as biased, subjective and anecdotal - especially if it challenges their own views. However, it can often provide insights or at least suggest alternative interpretations of data or events that are not available elsewhere. Careful consideration of such issues will be a useful starting point to determine your inclusion/exclusion criteria.

Inclusion/exclusion criteria

When programming search engines you can usually set inclusion and exclusion criteria to ensure you are not looking through material that you will not use. Taking time to set appropriate criteria will save you time in the long run, though it may be useful to do general quick search using Google to ensure you have not missed anything important by setting these restrictions.

Example: Selection criteria relevant to a health systems intervention in Ghana
Languages: English, French
Years: 2010-2015
Publications: Journals, books, dissertations, reports of specified agencies.
Regions: West Africa

Where to search

Spend a little time researching the most appropriate databases to use for your research topic. A list is provided in part 2 of this chapter. Depending on the time available, once you have used one database, try another and see if the same information is coming up. If it is, you can be more confident that your strategy is well-focused and that you are finding the relevant literature. If you only have time to use one search engine use Google or Google Scholar (depending on your inclusion criteria) as these search most widely. If you use these search engines you may need to limit the literature you search through, for example by only reading through the first ten pages of results.

Procedure

Firstly, search for the keywords you have selected and synonyms of those keywords in your chosen databases. If you are using the approach suggested earlier, you will be undertaking a context-setting search, one or two more specific searches relevant to your research and one very specific search. As you learn more about the topic, open up the search to wider material by adding words used often in the research (for example look at the keywords in the journal articles you are bringing up). You may also want to search for more papers from key authors and journals you find, making use of 'related articles' features and using the bibliographies of relevant research. It is useful to record your search in a table such as that shown below. This will assist you in assessing the extent to which you can feel confident that you have compiled the most important material and provide others with evidence of your methods you have adopted. Note that some databases allow you to maintain a record of past searches.

Table 2: Search links between availability of hospital performance data and utilization

Database	Date	Keywords	Total hits	Relevant Studies
Science Direct	02/09/2014	Countries (list with separator OR) AND Hospitals AND (HMIS OR HIS OR Terms related to performance) AND Utilization	103	19

3.   Quality-assessment of studies

The next step is to select the items identified in the search that you will use, given that there is not time for a systematic review of all of the evidence. This part of a literature search is key as it will ensure you spend your time effectively, and read in detail only the research that you will potentially be including. There are many ways of doing this, but one way of quickly

assessing studies and ensuring you select the most appropriate is to use an appropriate assessment tool that takes into account a range of factors. The aim of this procedure is to provide an indication of which studies should be seen as contributing most significantly and robustly to understanding this topic and it will also mean the evidence you present is responsibly and judiciously selected. Note that funding agencies place considerable emphasis on the need for robust evidence to informing policy and programming; including suspect or out-dated materials will not be helpful if you are seeking their support.

Quality assessment can be problematic. Katrak et al. (2004) identified a list of 121 different critical appraisal tools (e.g. Understanding Health Research 2016). They concluded that there is no 'gold standard' for appraising studies as there is a lack of information on the development and validity of these tools and only a few have been seriously evaluated. One interesting example is an approach adapted from a report prepared by the UK Department for International Development (DFID 2014). They suggest a two-part evidence assessment (single study and evidence body assessment), but here we focus on the first stage. Depending on the time available, you could simply use the general theory behind this approach without formally writing down the assessments. The procedure outlined below involves reading the abstract and methodology of each study as a basis for including or excluding them. More detail on the methods can be found in Chapter 2 of Aveyard (2007). Many search engines allow you to copy citations into a document as you proceed, such that by the end of this process you have your selected literature. If you have more time, and want to include more detail, a table such as that shown below can help you remember key aspects of each study and is a way to organise your results.

Table 3: Findings of a critical assessment process

Author/date	Related theme	Aim of paper	Type of information	Main findings	Strengths and limitations

Assessment of evidence strength

For each individual study, we can consider the research type, research design, and methodology to arrive at a quality assessment. Such a procedure can either be seen as a rough guide as you select material, or it can be undertaken more formally and the selection criteria described with multiple descriptive keys. For example, an assessment of (P&E; EXP;H) might mean that a study is highly relevant, primary and empirical, experimental and high quality. Table 4 provides one approach to classifying studies by type, table 5 lists questions allowing assessment of various quality dimensions and table 6 provides an aggregation index based on these dimensions.

Table 4: Classification of research studies by type

Research Type	Research Design
Primary and Empirical (P&E)	Non-Experimental (NEX)
	Quasi-Experimental (QEX)
	Experimental (EXP)
Secondary (S)	Systematic Review (SR)
	Non-Systematic Review (NSR)
Theoretical or Conceptual (TC)	N/A

Source: DFID 2014:9

Table 5: Principles for assessing the quality of individual studies

Principles of Quality	Associated Questions
Conceptual Framing	Does the study acknowledge existing research?
	Does the study construct a conceptual framework?
	Does the study pose a research question or outline a hypothesis?
Appropriateness and rigour	Does the study present or link to the raw data it analyses?
	What is the geography/context in which the study was conducted?
	Does the study declare sources of support/funding?
Appropriateness	Does the study identify a research design?
	Does the study identify a research method?
	Does the study demonstrate why the chosen design and method are well suited to the research question?
Cultural sensitivity	Does the study explicitly consider any context-specific cultural factors that may bias the analysis/findings?
Validity	To what extent does the study demonstrate measurement validity?
	To what extent is the study internally valid?
	To what extent is the study externally valid?
	To what extent is the study ecologically valid?
Reliability	To what extent are the measures used in the study stable?
	To what extent are the measures used in the study internally reliable?
	To what extent are the findings likely to be sensitive/changeable depending on the analytical technique used?
Cogency	Does the author 'signpost' the reader throughout?
	To what extent does the author consider the study's limitations and/or alternative interpretations of the analysis?
	Are the conclusions clearly based on the study's results?

Source: DFID 2014:14

Table 6: Study quality category definitions

Study Quality	Definition
High (H)	Demonstrates adherence to principles of appropriateness/rigour, validity and reliability; likely to demonstrate principles of conceptual framing, openness/transparency and cogency.
Moderate (M)	Some deficiencies in appropriateness/rigour, validity and/or reliability, or difficulty determining these; may or may not demonstrate principles of openness/transparency and cogency.
Low (L)	Major and/or numerous deficiencies in appropriateness/rigour, validity and reliability; may/may not demonstrate openness/transparency and cogency.

Source: DFID 2014a:15

4. How to synthesise your findings

The next stage is to summarise the findings of the literature search. This will provide readers with details as to your review methodology and findings. If you have broken your search up into the three areas suggested in section 1, and used a table as suggested in section 2 to note down key findings as you have been searching, this process should be fairly simple as you will have three tables summarising the key findings for the different sections of your search. The inverted triangle diagram could be used to structure your review. There are different approaches to this, and it partly depends on what you have been asked to do. You could

include several paragraphs on how you have conducted your search and use the inverted triangle diagram to summarise the results of the research. The aim is to interpret the results and consider the differences and similarities in different papers, rather than simply summarise them. This will give a new meaning to the results and identify gaps in the literature. These should be outlined to show how your research will add to the existing literature and why it is important to study this area.

If more detail is needed, a meta-ethnographic approach to synthesising information could be used. Developed by Noblit and Hare (1988), this approach involves determining keywords, phrases, metaphors and ideas that occur in some or all of the studies and interpreting these in the light of those identified in other studies (Britten et al. 2002). The aim of this is to determine the relationship between the studies so that consistencies and differences are identified. If further time was given to research or if the funder asks how you could expand your review, a meta-summary should be conducted, assigning codes to points discussed in each research paper and further sub-themes could be developed under each section (more detail can be found in Chapter 6 of Aveyard (2007)).

Finally, note that a narrative review such as this can lead to misleading conclusions and should be seen as a preliminary step towards undertaking the type of systematic literature review discussed in the second part of this chapter. It can be useful to clarify this at the end of your method statement and not interpret your findings too widely or make assertions that are not justified from the amount of time you have spent researching the issue. Do not be tempted to bend the data to show the gaps you would like them to show, to improve your argument or to align your review with stakeholder or donor perspective as this will cause problems later.

References

Aveyard, H. (2007). *Doing a Literature Review in Health and Social Care: A Practical Guide*, New York: Open University Press. www.chc.brookes.ac.uk/dr-helen-aveyard (accessed 12 March 2015)

Britten, Nicky, Rona Campbell, Catherine Pope, Jenny Donovan, Myfanwy Morgan, Roisin Pill. Using meta ethnography to synthesise qualitative research: a worked example. *Journal of Health Services Research Policy* 7(4):209-215. http://betterevaluation.org/sites/default/files/Using-meta-ethnography-to-synthesise-qualitatie-research.pdf

Cochrane A L (1972). *Effectiveness and efficiency: Random reflections on health services*. The Nuffield Provincial Hospitals Trust. http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/Effectiveness_and_Efficiency.pdf

DFID (2014). *How to Note on Assessing the Strength of Evidence*. www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf (accessed 12 March 2015)

DFID (2014a). Secure property rights and development: Economic growth and household welfare. *Property rights evidence paper*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/304551/Property-rights-evidence-paper.pdf (accessed 12 March 2015).

J-PAL (undated). *Introduction to Evaluations*. https://www.povertyactionlab.org/research-resources/introduction-evaluations

Katrak, P., Bialocerkowski, A.E., Massy-Westropp, N., Kumar, S. and Grimmer, K.A. (2004). A Systematic Review of the Content of Critical Appraisal Tools. *BMC Medical Research Methodology* 4(22)1-11. www.biomedcentral.com/1471-2288/4/22 (accessed 12 March 2015)

Lund, Hans, Klara Brunnhuber, Carsten Juhl, Karen Robinson, Marlies Leenaars, Bertil F Dorch, Gro Jamtvedt, Monica W Nortvedt, Robin Christensen, Iain Chalmers (2016). Towards evidence based research. *BMJ* 355:i5440 doi: 10.1136/bmj.i5440. http://www.bmj.com/content/355/bmj.i5440

NIH (2016). *Health Services Research and Health Policy Grey Literature Project: Summary Report*. https://www.nlm.nih.gov/nichsr/greylitreport_06.html#_Toc124061662

Noblit, G.W. and Hare, R.D. (1988). *Meta-Ethnography, Synthesising Qualitative Studies*. London: Sage Publications. https://uk.sagepub.com/en-gb/eur/meta-ethnography/book2416

NYU Libraries (undated). *Nursing Resources: A Self-Paced Tutorial and Refresher*: 1.7 Search Strategies: Boolean Logic. http://guides.nyu.edu/c.php?g=276860&p=1848399

Reid, M., Taylor, A., Turner, J. and Shahabudin, K. (undated). *Starting a Literature Review*. University of Reading. www.reading.ac.uk/web/FILES/sta/A5_Literature_Reviews_1_Starting.pdf (accessed 12 March 2015)

Understanding Health Research (2016). A tool for making sense of health studies. MRC/CSO Social and Public Health Sciences Unit, University of Glasgow. http://www.understandinghealthresearch.org/

A useful resource

The British Library for Development Studies provides a document delivery service, which can be useful if you do not have access to a particular article, http://blds.ids.ac.uk/index.cfm?objectid=D3FBAB71-4D85-11E0-A71C00016C1BDD3E (accessed 12 March 2015)

Part 2: Systematic reviews

Richard Longhurst
Institute of Development Studies

1.   Introduction

There has been an explosion in medical, nursing and allied health care professional publishing over the last 50 years. There are perhaps 20,000 journals and as many as two million articles per year. These keep expanding in number, making it literally impossible to keep up with primary research across the health domain. Even for specific research topics, the number of published studies can run to hundreds if not thousands. Some of these may give unclear, confusing or contradictory results, or involve research methods that are not compatible with those in other studies. There has also been a huge expansion in the number of such publications available via the internet, and researchers face the challenge of building skills that will enable them to use the electronic media in ways that allow effective access to this enormous volume of information. In addition, health care professionals have a wide range of information needs Akobeng (2005), requiring good quality information on the relevance, effectiveness, feasibility and appropriateness of a large number of health systems and services interventions.

Traditional reviews of the literature often lacked rigour because of the self-selection of research studies and subjective interpretation of the evidence. Recommendations based on such reviews would frequently be dismissed as biased. There was a turnaround in opinion in the 1980s-90s, with many arguing that traditional approaches had largely failed to extract useful and unbiased information. What was needed was the same rigour in secondary research (research where the objects of study are other research publications) as is expected from primary research i.e. original studies. Systematic Reviews (SRs) were designed to meet this challenge. They are based on an evidence translation mechanism undertaken in a highly rigorous, transparent and independent manner with full information on each stage of the procedure made available to the reader. They follow a strict peer-review protocol, with the reviewer starting the process with an open mind (NCCMT undated).

2.   Substance of a Systematic Review

A Systematic Review is a summary of existing research on a particular topic or research question. Although it is in essence a literature review it aims to use the same principles and rigour that is expected of primary research with generally accepted approaches and methods. This means that readers can be confident that common methods have been used that are well accepted and that comparisons can be legitimately drawn between SRs. The method involves interrogating multiple databases and search bibliographies for references, both published studies and also 'grey' material (unpublished but generally available material). SRs screen studies for relevance, appraise for quality on the basis of the research design, methods and the rigour with which each of these were applied, and synthesises the findings using predetermined formal quantitative or qualitative methods.

SRs are in a period of rapid development. In the health field, many still look at clinical- and cost-effectiveness, but methods now exist for reviewers to also examine issues of appropriateness and feasibility. Note that the use of the term Systematic Review does not guarantee the quality of the study. A number of apparent SRs have been published that fail to follow the prescribed protocols or adopt procedures that are likely to deliver biased findings. Each review needs to be interrogated by asking a series of questions that can uncover deficiencies (Shea et al 2009). As indicated in the introduction to this chapter, much of the

work on formalising the SR process was undertaken by the Cochran Collaboration Their handbook (Cochran Collaboration 2011) is regarded by many as the authoritative text on how to both conduct and report the findings of SRs.

3.   Steps in a Systematic Review

There are eight main steps to a SR process (Mann and Weightman 2015). The first is to identify a health care question clearly and unambiguously. Generally SRs answer specific healthcare questions and assess the effectiveness of particular interventions rather than providing general summaries of the literature on a given topic. With the example of an intervention, the review question would clearly define: the specific population or problem being investigated, the intervention being evaluated, the comparison or control under investigation and the outcome of interest.

Second, a review protocol is developed. This is a detailed description of the scope, aims and methods of the study, stating clearly the review question, how and where studies will be located, selected, appraised and synthesised. This allows any problems of bias to be addressed. In recent years, those undertaking SRs have been encouraged to include their protocols in a central database called PROSPERO. This database can be searched to locate existing SR protocols relating to specific types of intervention.

The third step is the search of the literature with the aim of identifying all relevant studies on the research topic. You may start by using a general search engines such as Google Scholar, talking to experts in the field, and looking at book reviews. This will guide the design of the comprehensive search strategy required for a SR, for example by identifying the most important journals and keywords. This search strategy must be clearly specified in the review protocol. For a health systems intervention the list of databases searched can be very extensive, as shown in table 1. Note that many of these only provide services on payment of a subscription fee, so you will need to check if your institution has access. If not, many of the same journals may be available via the HINARI initiative of the WHO.

Table 1: Detailed list of databases searched for a Systematic Review of the Integrated Management of Childhood Illness (IMCI) strategy	
Database	Free (F) or Subscription (S)
Cochrane Central Register of Controlled Trials	F
MEDLINE/PubMed	F
EMBASE	S
Cumulative Index to Nursing and Allied Health Literature (CINAHL)	S
Latin American Caribbean Health Sciences Literature (LILACS)	F
WHO Library & Information Networks for Knowledge Database (WHOLIS)	F
Science and Social Sciences Citation Indices (Web of Science)	S
Population Information Online (POPLINE)	F
WHO International Clinical Trials Registry Platform (WHO ICTRP)	F
Global Health	S
Ovid	S
Scopus	S
Proquest Health Management Database	S
Proquest Public Health Database	S

Source: adapted from Gera et al. (2016)

In all fields there is a tendency to publish research with positive findings: research that shows 'no effect' may not be published but is just as important in terms of gaining an overall picture

of the effect of an intervention. This 'publication bias' should be addressed by seeking out unpublished studies, which, as indicated above, are generally described as the 'grey literature' (Gray 1998). However, finding unpublished work can be very difficult because of the lack of a public record. A major initiative in this area is GreyNet International. This is a subscription based organisation but its website also provides links to a number of open access sources. It is also possible to search databases of conference proceedings (NIH 2016), higher degree dissertations (OATD undated), reports from international (e.g. WHO and UNICEF) and national donor agencies and the websites of selected schools of public health. In addition English language 'bias' should be addressed. If other languages are generally excluded (due to a lack of resources for translation), this should at least be noted, and the option of identifying and translating a small number of key articles considered. If possible, the search results should be imported into reference management databases such as Endnote or the freeware alternative, Zotero.

The fourth step is to identify relevant studies. In a formal SR, studies are assessed for their actual relevance independently by two or more researchers. The criteria for inclusion (i.e. which population, intervention and outcome measures are of interest) should be documented in the review protocol. Pre-specifying inclusion and exclusion criteria protects the review from allegations of investigator bias, where the reviewer for one reason or another becomes attached to one line of reasoning and tends to selects studies which confirm that option.

A key eligibility criterion relates to the type of research design adopted by the study. SRs were initially used for reporting on clinical trials, where double-blinded randomised controlled trials (DBRCTs), were regarded as the 'gold standard' in terms of reducing the possibility of biased findings. However, with their application to general health sector interventions, where double-blinding is typically impossible and RCTs often not feasible, it has become common to include a wide variety of experimental and non-experimental designs. For example, one recently submitted protocol for a review of economic evaluations of m-Health interventions specifies the inclusion criteria as:

"*Randomised controlled trials (RCTs), quasi-RCTs, controlled clinical trials (CCT), controlled before-after-studies (CBA), interrupted time series (ITS) and before-after or cohort type evaluations, undertaken with formal health economic evaluations (cost-effectiveness analysis (CEA), cost-benefit analysis (CBA), cost-minimization analysis, cost-consequence analysis, and cost-utility analysis). Economic modelling studies will also be considered. Published in the English language*" (Iribarren 2014).

There will often be a trade-off between preferred research designs and the number of studies included in the review. This is well illustrated in a Cochran SR of interventions intended to reduce corruption in the health sector. The eligibility criteria in terms of research designs were described for two types of analysis:

"*For the primary analysis, we included randomised trials, non-randomised trials, interrupted time series studies and controlled before-after studies that evaluated the effects of an intervention to reduce corruption in the health sector. For the secondary analysis, we included case studies that clearly described an intervention to reduce corruption in the health sector, addressed either our primary or secondary objective, and stated the methods that the study authors used to collect and analyse data*". (Gaitonde et al. 2016)

In the event, no studies were found that met the criteria specified for the primary analysis, while nine were accepted for the secondary. It can often be useful to categorise studies in this way and then consider how much weight to give to findings from the various types of design. A guide to research designs and their strengths and limitations in terms of potential bias can

be found in Chapter 13 of the Cochran Handbook (Cochran Collaboration 2011), and discussion of designs for public health interventions at paragraph 21.2 of that volume.

Step five is to critically appraise those relevant studies. As above, in a formal SR process it is strongly recommended that the appraisal should be performed independently by two or more researchers to avoid bias. The appraisal centres on the methodology adopted and the rigour with which the research appears to have been conducted, based on the published report. The appraisal will typically be undertaken using a formal checklist. These will vary depending on the type of study (SURE undated), in particular they will be very different for experimental (SURE 2015), observational studies (CASP 2013a) and qualitative studies (CASP 2013b).

Step six is the extraction of findings to construct a table allowing direct comparison of the main findings from each study. This is a difficult phase of the SR and one at which considerable judgement needs to be applied. It is complicated by issues such as incomplete reporting of study findings, the large range of outcomes commonly used to evaluate an intervention and the different ways in which data are reported and presented. Table 2 illustrates how such a table was constructed for a review of studies on the effects of mHealth interventions for chronic illnesses. Each blank cell indicates that a finding was not reported for the related study.

Table 2: Comparison of findings from studies of mHealth interventions

	Balsa and Gandelman	Shetty et al.	Shahid et al.	Ostojic et al.
Intervention	Health promotion & awareness		Remote monitoring & care support	
Design	RCT	RCT	RCT	RCT
Condition	Diabetes	Diabetes	Asthma	Asthma
Intervention group	195	110	220	8
Control group	193	105	220	8
Clinical outcomes				
Blood pressure	+/-		++	
HbA1c		++	++	
Coughing				++
Compliance outcomes				
Adherence to diet		+/-	++	
Adherence to exercise		+	++	
Knowledge	+/-			

Notes: +/- no difference between intervention and control groups
     + non-significant positive difference between intervention and control groups
     ++ significant positive difference between intervention and control groups
Source: adapted from Stephani et al. (2016):p7

The seventh step is to summarise the conclusions of the studies. The aim is to synthesise the individual studies to provide a clear and unambiguous judgment on the effectiveness of the intervention and a systematic summary of the research studies. In clinical studies, where a number of studies typically address precisely the same question, use similar populations, administer the intervention in the same manner and measure identical outcomes, it is often possible to combine the data statistically in a meta-analysis (Haidich 2010) to get an overall estimate of the effectiveness of an intervention. However this approach will rarely be appropriate for health systems interventions.

The results can often be reduced to a simple categorisation of studies that showed the specific intervention was beneficial, and those that indicated that it was not. A synthesis may also be achieved by a narrative summary supported by brief descriptions of each study in 'evidence tables' (Spiva 2013). Bodies of evidence should be summarised in terms of four characteristics (DFID 2014): i) the technical quality of the studies constituting the body of evidence and the

degree to which risk of bias has been addressed; ii) the size of the body of evidence; iii) the context in which the evidence is set; and iv) the consistency of the findings produced by studies constituting the body of evidence.

The final step is to document the review findings. SRs need to be promoted to inform policymakers and practitioners and so are useless unless they help fuel this objective. Report production and dissemination are crucial parts of the process, written along a focussed structure of introduction, methodology; nature of evidence identified and detailed findings, conclusions and recommendations. There needs to be a clear description of the methods so the reader can judge the validity of the techniques employed.

SRs do have some drawbacks. When well conducted they should give the best possible estimate of any true effect but such confidence may be misplaced on some occasions. First, SRs may simply be badly done. A checklist, such as that indicated below, can be used to determine the level of quality. Second, there may be inappropriate aggregation of studies that differ in terms of the nature of the intervention, the target population or types of data gathered that can lead to the drowning of important effects. For example, the effects seen in some subgroups may be concealed by a lack of effect (or even contrary effect) in other subgroups. Finally, when the findings of SRs are not in harmony with the findings from large scale single research exercises, they need to be weighed against potentially conflicting evidence from other sources.

4.   An Appraisal Framework for SRs

Akobeng (2005) suggests that some of the key questions to be addressed in relation to any systematic review are:

- Did the review address a clearly focused question?
- Did the review include the right type of study?
- Did the reviewers try to identify all relevant studies?
- Did the reviewers assess the quality of all the studies included?
- If the results of the studies have been combined, was it reasonable to do so?
- How are the results presented, and what are the main results?
- How precise are the results?
- Can the results be applied to your local population?
- Were all important outcomes considered?
- Should practice or policy change as a result of the evidence contained in this review?

5.   General Issues and the Future

The key element of SRs is impartiality, hence the requirement for independent assessment. However they are not easy, requiring enormous care and rigour with considerable attention to methodological detail and analysis. The label of 'systematic review' is hard earned. There are some changing trends in SRs. Increasingly health professionals cannot wait for a year or so for a full SR to produce its findings. Rapid Evidence Assessments (REAs), or Rapid Reviews (RRs) (Polisensa 2015) can provide what is already known about a topic or intervention, and take about two to six months. They use systematic review methods to search and evaluate the literature, but the comprehensiveness of the stages may be limited. The use of these approaches depends on the time frame for decisions, uncertainty about effectiveness when there has already been considerable prior research or to develop a map of evidence to determine the existing evidence and to direct future research needs.

References

Akobeng, A K (2005). Understanding systematic reviews and meta-analysis. Arch Dis Child 90:845-848. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1720526/pdf/v090p00845.pdf

Baker P, J Costello, M Dobbins, E Waters (2014). Cochrane Update: The benefits and challenges of conducting an overview of the benefits of public health: a focus on physical activity. *Journal of Public Health* 36 (3):517-521. http://eprints.port.ac.uk/16776/1/The_benefits_and_challenges_of_conducting_an_overview_.pdf

Banberry A, A Roots and S Nancarrow (2014). Rapid Review of applications of e-health and remote monitoring for rural residents. Australian Journal of Rural Health 22:211-222. https://www.researchgate.net/publication/266745913_Rapid_review_of_applications_of_e-health_and_remote_monitoring_for_rural_residents

Boonstra A, A Verslius and J Vos (2014). Implementing Electronic Health records in Hospitals: A systematic literature review. *BMC Health Services Research* 14:370. https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-14-370

Burlan D, K Buse, J Shiffman and S Tanaka (2014). The bit in the middle: a synthesis of global health literature and policy formulation and adoption. *Health Policy and Planning* 29: iii23-iii34. http://heapol.oxfordjournals.org/content/29/suppl_3/iii23.full.pdf+html

CASP (2013a). *Case Control Study Checklist*. Critical Appraisal Skills Programme. http://media.wix.com/ugd/dded87_19dd1d558a9977c0e0b30cedf86a9da7.pdf

CASP (2013b). *Qualitative Research Checklist*. Critical Appraisal Skills Programme. http://media.wix.com/ugd/dded87_29c5b002d99342f788c6ac670e49f274.pdf

Cochrane, A L (1972). *Effectiveness and Efficiency: Random Reflections on Health Services*. The Nuffield Provincial Hospitals Trust. www.nuffieldtrust.org.uk/sites/files/nuffield/publication/Effectiveness_and_Efficiency.pdf

DFID (2014). Assessing the Strength of Evidence. How To Note. DFID. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf

Erasmus, E, M Orgill, H Schneider, and L Gilson (2014). Mapping the existing body of health implementation research in lower income settings, what is covered and what are the gaps? Health Policy and Planning 29: iii35-iii50. http://heapol.oxfordjournals.org/content/29/suppl_3/iii35.full.pdf+html

Gaitonde, Rakhal, Andrew D Oxman, Peter O Okebukola, Gabriel Rada (2016). Interventions to reduce corruption in the health sector. *Cochran Database of Systematic Reviews*. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008856.pub2/abstract

Gera, Tarun, Dheeraj Shah, Paul Garner, Marty Richardson, Harshpal S Sachdev (2016). Integrated management of childhood illness (IMCI) strategy for children under five. *Cochrane Database of Systematic Reviews*. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010123.pub2/abstract;jsessionid=DD6933A2A49AF9A93AFEF43025B18B30.f04t02

Gilson L (2014). Qualitative Research Synthesis for research policy analysis: what does it entail and what does it offer? *Health Policy and Planning* 29:iii1-iii5. http://heapol.oxfordjournals.org/content/early/2014/11/11/heapol.czu121

Gilson, L, H Schneider and M Orgill (2014). Practice and Power: a review and interpretive synthesis focused on the exercise of discretionary power in policy implementation by front-line providers and managers. Health Policy and Planning 29: iii51-iii69. http://heapol.oxfordjournals.org/content/29/suppl_3/iii51.full.pdf+html

Gray, Bradford H. (1998). Sources used in health policy research and implications for information retrieval systems. *Journal of Urban Health* 75(4): 842-852. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3455996/pdf/11524_2006_Article_BF02344512.pdf

Greenhalgh T, G Robert, F MacFarlane, P Bate, O Kyriakidou and R Peacock (2005). Storylines of research in diffusion of innovation, a meta-narrative approach to systematic review. *Social Science and Medicine* 61:417-430. http://www.sciencedirect.com/science/article/pii/S0277953604006471

Greenhalgh T (2014). Evidence based medicine: a movement in crisis? *BMJ* 348:g3725. http://www.bmj.com/content/348/bmj.g3725

Heidich, AB (2010). Meta-analysis in medical research. *Hippokratia* 14(Suppl 1): 29-37. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049418/pdf/hippokratia-14-29.pdf

Iribarren, Sarah (2014). A systematic review of economic evaluations of mHealth solutions at large in low resource settings/low- and middle-income countries. *PROSPERO International prospective register of systematic reviews*. Centre for Reviews and Dissemination. University of York. http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014014913

Jagosh J, Macaulay AC, Pluye P, Salsberg J, Bush PL, Henderson J, Sirett E, Wong G, Cargo M, Herbert CP, Seifer SD, Green LW, Greenhalgh T (2012). Uncovering the benefits of participatory research: Implications for a realist review of health research and practice. *Millbank Quarterly* 90(2):311-346. https://depts.washington.edu/ccph/pdf_files/Jagosh_etal-2012-Milbank-Uncovering-the-Benefits-of-PR-FINAL.pdf

Li, Xiuxia, Zheng Ya, Chen Yaolong, Yang Kehu, and Zhang Zongdiu (2015). The reporting characteristics and methodological quality of Cochrane reviews about health policy research. *Health Policy* 119:503-510. http://www.healthpolicyjrnl.com/article/S0168-8510(14)00232-2/pdf

Loevinsohn, M, L Mehta, K Cuming, A Nicol, O Cuming and J Ensink (2015). The cost of a knowledge silo, a systematic re-review of water, sanitation and hygiene interventions. *Health Policy and Planning* 30 (5): 660-674. http://heapol.oxfordjournals.org/content/30/5/660

Mann, Mala and Alison Weightman (2015). Beginning a systematic review of the health care literature. http://www.cardiff.ac.uk/insrv/resources/guides/inf113.pdf Accessed 15 March 2015.

NIH (2016). US National Library of Medicine: Conference Proceedings. https://www.nlm.nih.gov/services/Subject_Guides/conferenceproceedings/

OATD (undated). Open Access Theses and Dissertations. https://oatd.org/

Pawson, R, J Greenhalgh, C Brennan and E Glidewell (2014). Do reviews of health care interventions teach us how to improve health care systems? *Social Science and Medicine* 114:129-137. http://eprints.whiterose.ac.uk/79346/1/Pawson%20%282014%29.pdf

Polisena, Julie, Chantelle Garritty, Chris Kamel, Adrienne Stevens and Ahmed M Abou-Setta (2015). Rapid review programs to support health care and policy decision making: a descriptive analysis of processes and methods. *Systematic Reviews* 4(26):1-7. http://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-015-0022-6

Rockers P, J-A Rottingen, I Shemilt, P Tugwell, and T Barninghausen (2015). Inclusion of quasi- experimental studies in health systems research. *Health Policy* 119: 511-521. https://www.researchgate.net/publication/273703703_Inclusion_of_quasi-experimental_studies_in_systematic_reviews_of_health_systems_research

Shea, Beverly J., Candyce Hamel, George A. Wells, Lex M. Bouter, Elizabeth Kristjansson, Jeremy Grimshaw, David A. Henry, Maarten Boers (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology 62:1013-1020. http://www.ciet.org/_documents/publication%20-%20amstar%20is%20reliable.pdf

Spiva, LeeAnna (2013). *Building your evidence table*. Wellstar. http://www.wellstar.org/education/documents/nursingresearchconference/2013/2013-spiva-evidence-table.pdf

Stepani, Victor, Daniel Opoku and Wilm Quentin (2016). A systematic review of randomized controlled trials of mHealth interventions against non-communicable diseases in developing countries. *BMC Public Health* 16:572. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4946127/pdf/12889_2016_Article_3226.pdf

SURE (2015). *Questions to assist with the critical appraisal of randomised controlled trials and other experimental studies*. Specialist Unit for Review Evidence. Cardiff University. http://www.cardiff.ac.uk/__data/assets/pdf_file/0006/212766/SURE_RCTs-and-other-experimental-studies_Checklist_2015-update.pdf

Thomas, J, A O'Mara-Eves, and G Brunton (2014). Using Qualitative Comparative Analysis (QCA) in systematic reviews of complex interventions: a worked example. *Systematic Reviews* 3:67. https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/2046-4053-3-67

University of Reading (undated). *Starting a Literature Review*. Study Advice and Maths Support. http://www.reading.ac.uk/web/FILES/sta/A5_Literature_Reviews_1_Starting.pdf

Walt G and L Gilson (2014). Can frameworks inform knowledge about health policy processes? Reviewing health policy papers on agenda setting and testing them against a specific health priorities framework. *Health Policy and Planning* 29: iii6-iii22. http://heapol.oxfordjournals.org/content/29/suppl_3/iii6.full.pdf+html

Witter, S, A Frethiem, F Kessy and A Lindahl (2012). *Paying for Performance to improve the delivery of health interventions in low- and middle-income countries*. Cochrane collaboration. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007899/pdf

Wong, G, T Greenhalgh, G Westrop, G Buckingham, and R Pawson (2013). RAMESE publication standards: realist synthesis. *BMC Medicine* 11: 21. https://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-11-21

Yoong S, T Clinton McHarg, L Wolfendon (2015). Systematic Reviews examining implementation of research into practice and impact on population health are needed. *Journal of Clinical Epidemiology* 68(7):788-791 http://dx.doi.org/10.1016/j.jclinepi.2014.12.008 Accessed March 12 2015.

Some useful resources:

The Cochrane Library www.cochrane.org
The Joanna Briggs Institute www.joannabriggs.edu.au/pubs/systematic_reviews.php
The Campbell Collaboration www.campbellcollaboration.org
The Centre for Evidence-Based Medicine www.cebm.net
The NHS Centre for Reviews and Dissemination www.york.ac.uk/inst/crd
Bandolier www.medicine.ox.ac.uk/bandolier
Network of African Medical Librarians http://karibouconnections.net/medlibafrica/#cour

Chapter 4: Understanding the Intervention

Henry Lucas and Gerald Bloom
Institute of Development Studies

It is a curious fact that many implementation plans contain very limited information as to precisely how the intervention with which they are concerned will produce the expected results. They often provide very detailed accounts of the various inputs required and activities to be undertaken, together with an impressive list of potential outcomes. But many have what has been described as a "missing middle" (Lucas et al. 2004:21). They do not spell out the detailed process whereby the identified activities can be expected to achieve the intended goals.

For policy-driven interventions, there may be ideological reasons for such omissions. For example, if there is a conviction among policy makers that 'pay for performance' is an obvious way to improve services, they may not question too closely the precise mechanics of an intervention based on this approach. Where interventions are funded by an external donor, those seeking that funding may well focus their attention on ensuring that the discussion of the intervention and associated outcomes will appeal to that donor and spend less time on setting out the 'fine details' of their implementation plan. It may also be that some implementers sincerely believe that such details will almost certainly prove to be irrelevant when the implementation moves from the design phase to confront the complexities of the real world. This attitude may be seen as similar to that of Helmuth von Moltke, head of the Prussian army in the 19th century, when he suggested that "*No battle plan ever survives first contact with the enemy*".

An alternative view, also proposed by another war-time leader, Dwight D Eisenhower, was that "*Plans are worthless but planning is everything*". This latter sentiment seems much more useful from an implementation research perspective. It suggests that the more understanding you have of the implementation plan, including the underlying assumptions and potential risks, the more rapidly can you become aware of explicit or implicit modifications to that plan, whether these are driven by changing attitudes within the implementation team or by the external context.

1. The Implementation Dilemma

Interventions (policy changes, projects, programmes) can be seen as attempts to transfer health innovations that have demonstrated efficacy in the laboratory, clinical trials or small-scale pilot studies to benefit larger populations. Those involved in the development of such innovations are often very concerned that they should not be modified in ways that they fear may reduce efficacy. They focus on the issue of *Fidelity* (Carroll et al. 2007, Perez et al. 2011). On the other hand, local health experts will be primarily concerned with the effectiveness of the specific implementation that will affect the lives of the population they serve. They would see the potential for success of that implementation as being greatly enhanced by appropriate *Adaptation* to the local context.

There is therefore a basic dilemma which confronts all those who design the implementation of a promising intervention:

- The more rigidly implementation is controlled to ensure fidelity to the intervention, the more likely it will be that local factors (resource constraints, inadequate infrastructure, cultural factors, etc.) will reduce effectiveness.

- The more an implementation is adapted to local conditions, the more difficult it will be to argue that findings can be generalised to other localities or populations.

An additional problem faced by those who seek to promote the use of evidence-based interventions is that there may be considerable uncertainty as to the extent to which the intended intervention has been modified. The claimed degree of fidelity may be substantially less than the actual, as implementers make perfect sensible but often undocumented adjustments to overcome local barriers or bottlenecks.

Clinic research compared to implementation research

There is a long history of clinical trials that are often very convincing in terms of: 'what works?' or even 'what works best?' Experience with implementation research has been less encouraging - there are many interesting individual studies but limited accumulated knowledge that can be applied to new interventions. The underlying problem is that simple technical interventions typically involve complex social interventions that result in:

- Context dependency - low fidelity
- Outcomes that depend on detailed processes and pathways that are often not well understood.

In many cases it would seem that '*the intervention is the implementation*', i.e. we have a series of both successes and failures for what the implementers describe as the 'same' intervention but where there have in fact been considerable adaptations to local contexts, sometimes discarding or radically amending what might be regarded as essential features of the intended intervention.

Implications for IR

The above implies a need to seek an in-depth understanding of: (1) the intervention (for example identifying those elements seen as essential and those which could be modified without undermining the intervention objectives); and (2) the planned implementation process, with particular attention to modifications driven by a perceived need for adaption to a specific local context. This suggests a need for a monitoring system that can track changes in the implementation process where there are any deviations from the original plan. A useful starting point is to construct (or review if one already exists) a 'logical model' for the intervention ([DFID 2011](#), [W.K. Kellogg Foundation 2004](#), [Gasper 1997](#)). Such models are commonly required by international donors to provide a simplified explanation as to how a specific intervention is intended to further their objectives.

2. The Logical Model

The logical model of an intervention is of the 'if-then' type, linking what are seen as the activities, outputs, purpose and goal of an intervention:

- If activities are undertaken then outputs should be produced.
- If outputs are produced then outcomes that serve the purpose should result.
- If outcomes result then they should contribute towards achieving the goal.

Note that those managing an intervention are seen as responsible for producing a defined and quantified set of 'outputs', as illustrated in figure 1. The output-to-purpose and purpose to goal steps rely on the validity of assumptions, based on existing evidence, made by those designing the intervention. Clearly, for each 'link in the chain' to function, a series of additional assumptions relating to the external context must also hold. The more certain we can be of the resilience of each of the links to changes in those external contexts, the more persuaded we will be as to the likely success of the implementation.

Figure 1: the components of a logical model
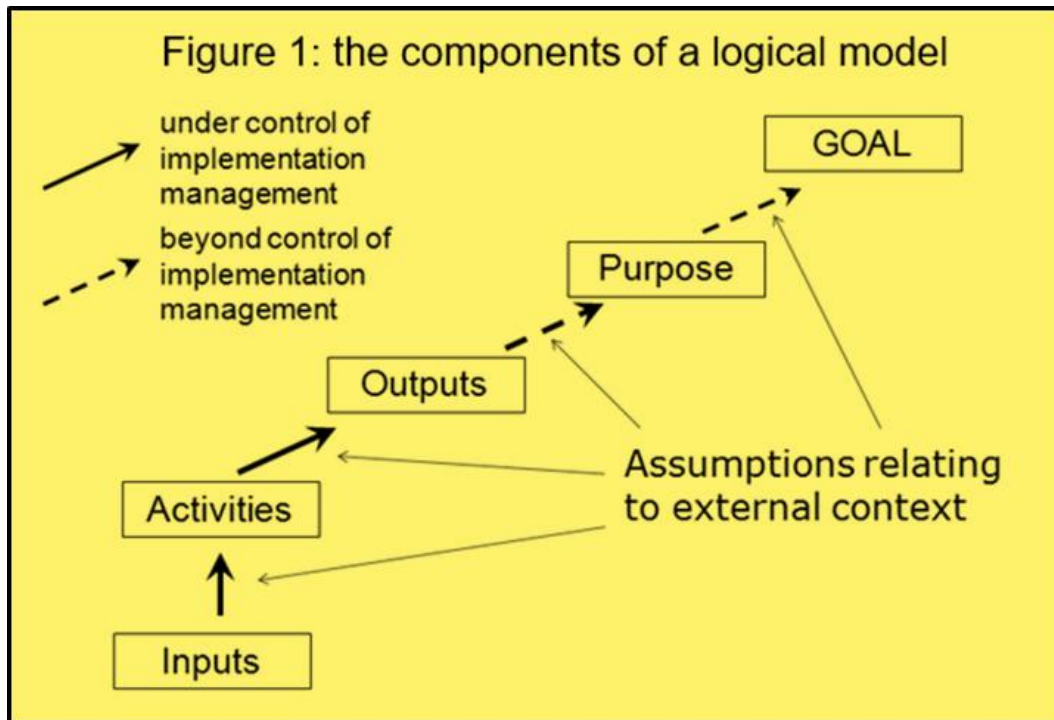
Logical Framework Analysis

A procedure based on the above, called Logical Framework Analysis (LFA), was originally introduced by international donor agencies as a management tool, designed to increase accountability and central control by imposing "*hierarchically ordered and quantified objectives*" (Gasper, 1997:3). Those objectives are often explicitly expressed as targets such as '70% of children immunised', '80% of households with an insecticide treated bed net'. It was seen, particularly by the recipients of funds, as a mechanism whereby donors assessed cost-effectiveness using Objectively Verifiable Indicators (OVIs) that were designed to allow progress monitoring and evaluation. The framework can be set out in simple matrix format.

Table 1: The Logical Framework Matrix

Vertical Logic	Objectively Verifiable Indicators (OVI)	Means of Verification (MOV)	Assumptions
GOAL			
PURPOSE			
OUTPUTS			
ACTIVITIES	Inputs		

As indicated above, the 'Vertical Logic' of the matrix links activities and outputs, which the implementation team has contracted to deliver, with the purpose and goal of the intervention as agreed between that team and those providing the necessary resources. The four levels are defined as follows:

- *Goal*: The higher level objective towards which the intervention is expected to contribute (e.g. reduced IMR).
- *Purpose*: Outcomes expected to be achieved as the result of the intervention (e.g. increased child immunization rates).

- **Outputs**: Results for which the implementation management are responsible (e.g. improved access to immunization).
- **Activities**: The activities that will be undertaken in order to produce outputs (e.g. reform of provider incentives).

Objectively Verifiable Indicators (OVI)

One primary purpose of the Logical Framework from an implementation research perspective is to raise questions as to how the key implementation inputs, outputs and outcomes can be effectively monitored to assess the extent to which implementation is progressing as intended and generating the expected outcomes. The framework requires the identification of a set of objectively verifiable indicators at each level:

- Goal: Measures to verify to what extent goals are fulfilled.
- Purpose: Measures to verify extent to which outcome targets are achieved.
- Outputs: Measures to verify extent to which output targets are achieved.
- Activities (Inputs): Measures of inputs (resources) used to undertake the activities.

Assumptions

The framework also requires identification of important conditions or events outside the control of the implementation management that are seen as necessary:

- To contribute to the goal.
- For the achievement of the purpose.
- For the production of outputs.
- For the implementation to start.

Assumptions are of particular interest for implementation research because of their relevance in assessment of the possibilities for scaling up or relocating the intervention. Some key questions to be addressed would be:

- Are the stated assumptions plausible in the existing context and how specific are they to that context?
- Are there important implicit (unidentified) assumptions?
- What consequences might flow from an incorrect assumption?
- As the implementation proceeds, have any assumptions proved to be incorrect?

Possible uses of the logical framework matrix

In its simplest form, as illustrated by the example in in table 2, the logical framework matrix can be seen as a brief summary of the basic underlying logic of an implementation and should allow an initial assessment of its plausibility and the extent to which it is context dependent. Perhaps one of its most valuable uses to permit discussion between the members of the implementation team and selected stakeholders to promote a common understanding as to how the implementation is expected to deliver the intended outcomes. One important task of the implementation researcher is to question both the extent to which the logic is plausible and the degree to which it depends on the contextual assumptions. In particular, the implementation researcher should attempt to identify those assumptions which can be seen as potentially determining the relative success or failure of the intervention. This can provide a useful guide to key issues that will need to be addressed in exploring possibilities for scaling up or re-location.

The framework can also be used to jointly identify and agree potential process, output and outcome indicators that can be used to verify if the implementation is proceeding as planned and producing the expected results. This implies a simultaneous process of identifying sources of data that can be used to determine those indicators. Together these activities should

provide a sound basis for the design of the implementation monitoring and evaluation system. Finally, if this should indicate that progress is not being made as intended, the logical frame matrix can be revised to reflect any necessary modifications to the original plan required to get the implementation back on track.

Table 2: Outline example of a Logical Framework Matrix

	Objectively Verifiable Indicators	Means of Verification	Assumptions
Goal: Reduced child malaria deaths	IMR, U5MR	Demographic and Health Surveys	Bed nets effective in preventing malaria in infants and young children
Purpose: Increased proportion of children under 5 sleeping under an insecticide treated bed net (ITN).	Proportion of children under 5 sleeping under an ITN	Demographic and Health Surveys	1. Communities persuaded of benefits. 2. No serious barriers to use within communities.
Outputs: 1. Mass distribution of ITNs 2. System to maintain & extend coverage 3. Communications activities to encourage appropriate use	ITNs distributed Distribution mechanisms established Number and types of communication activities	Implementation M&E system	1. Managers, providers released to attend training 2. No major delays in procurement process
Activities: 1.Procurement and distribution of ITNs 2. Establishment of sustainable supply system 3. Training for managers and providers 4. Production of communications materials 5. Communications activities 6. Development of reliable information systems	Inputs	M&E system Accounting systems	1. Support from local officials, providers, communities. 2. Financial resources provided on time

Criticism of Logical Frame Analysis has largely centred on its alleged rigidity in the context of what may well be a rapidly changing environment and for what is often characterised as a narrow and simplistic approach to interventions, particular when the frameworks are used to set often arbitrary targets. As indicated above, one response has been to emphasise the adoption of so-called 'process' Logical Frameworks, which can be modified during

implementation. The main concern here is that the ordered world of the logical framework, which may indeed be useful for the limited task of clarifying inputs, outputs, objectives and aspirations, should not be confused with the much more complex, highly politicised and extremely fluid environment which characterises many health system interventions.

3.   Logical Models and Theories of Changes

The basic logical framework matrix identifies intervention components and provides a useful summary of the 'chain of causality' linking inputs, activities and outcomes. A Theories of Change (ToC) model attempts to explain in much greater detail *how* the 'links in the chain' are intended to function - to develop 'an implementation theory'. (Vogel 2012, Mackenzie and Blamey 2005, Grantcraft 2006, International Network on Strategic Philanthropy 2005). Thus in the above example it was indicated that 'Communications Activities' would be used as a means to promote appropriate use of ITNs within target communities. A ToC model of this link might suggest, based on previous experience, that engaging with key opinion leaders and advertising in the local media prior to an open community meeting was likely to produce the desired outcome.

What is a Theory of Change?
The theory of change for an individual component of an implementation can be seen as a detailed flow diagram, as illustrated in figure 2, setting out the sequential processes required to achieve a given (intermediate or final) objective.

Figure 2: ToC for a behavioural change intervention
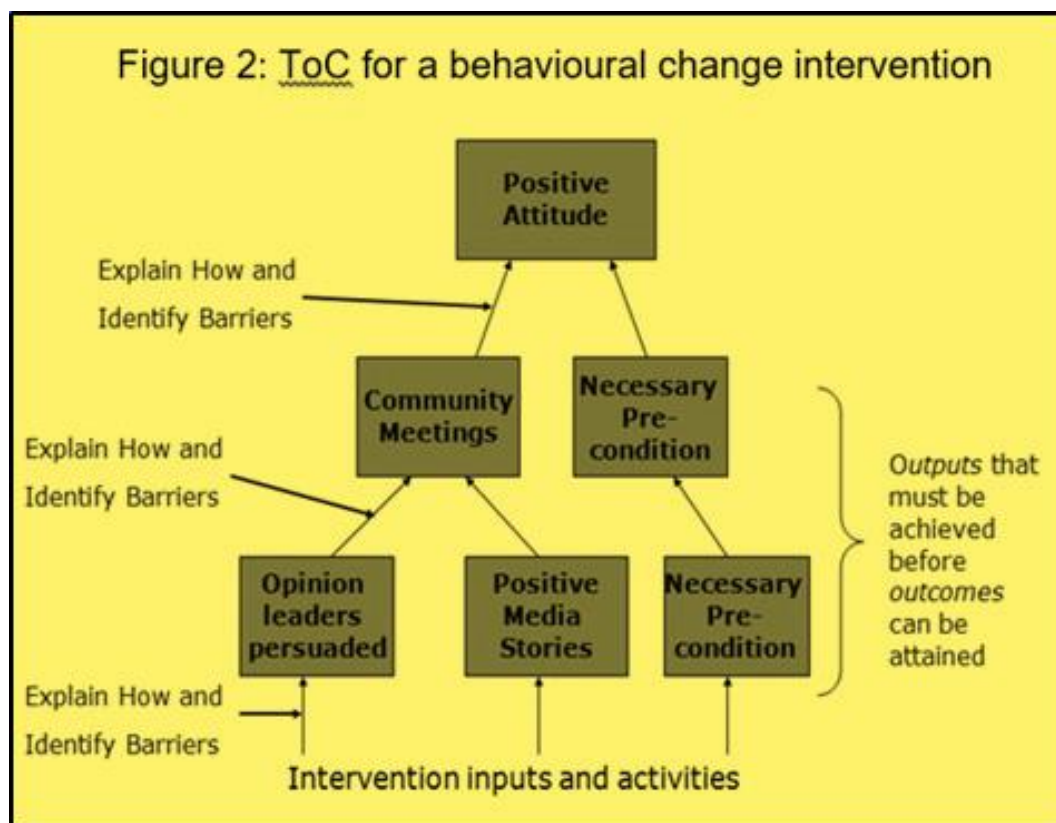
Testing ToC Models
ToC models should also specify quantified indicators that can be used to test the implementation theories. For example, in the above, it may be suggested that a majority of opinion leaders will have to be persuaded or that 75% of a community must be reached by the media publicity in order that the expected outcome will be achieved. Developing such ToC

theories is a much more difficult and time consuming than constructing a Logical Framework. For example, figure 3 below shows the detailed ToC for a prevention and treatment program for patients with heart disease (Mackenzie and Blamey (2005):158). It should be kept in mind that ideally such a ToC should be constructed in collaboration with representatives of all key stakeholders, which implies an extended and possibly contentious participatory process. The assumption is that by devoting sufficient resources to developing a shared understanding as to how an intervention is intended to work, we can design monitoring systems that will allow us to both test those theories and gain the knowledge that will allow us to adapt the implementation design if they prove incorrect.

Figure 3: Have a heart project – Theory of Change
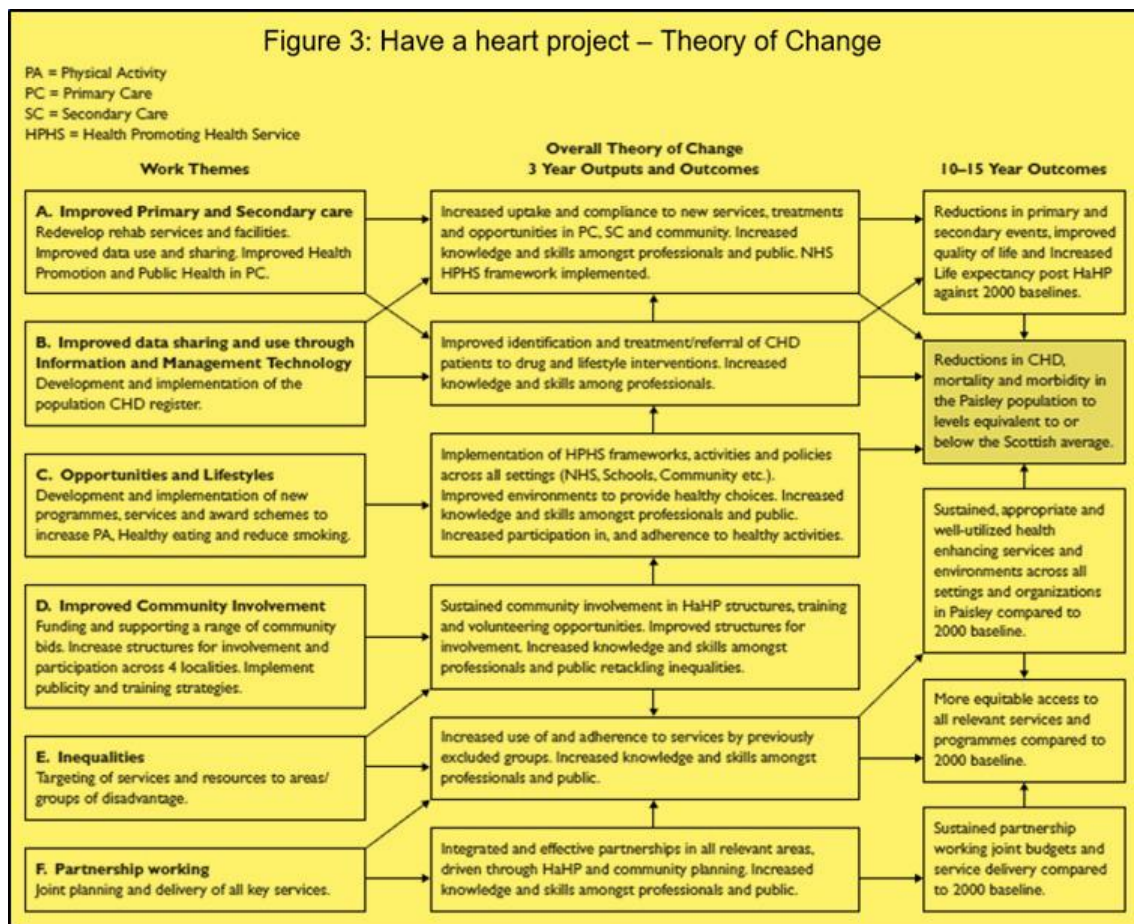
4.   Complex Adaptive Systems

Theories of change can be extremely complicated if an intervention contains multiple components. However, as with the Logical Framework, there is an underlying assumption that if the inputs and activities identified in the ToC can be implemented as intended, they will result in the desired outputs and outcomes. Comparison is often made to the first moon landing. That was one of the most complicated projects ever attempted but successful because a myriad of component parts functioned as planned, resulting in the predicted outcome. This feat of engineering can be contrasted, for example, with many biological systems, which are inherently unpredictable. We cannot know precisely when a seed will germinate, where a tree will form its first branch or which genes a child will inherit from each parent. Similar considerations can be applied to social systems. For example, it seems impossible to determine in advance which individual will emerge as the most influential in a political party, which children starting at a new school will become close friends or which marriages will be

successful. The notions of predictability and unpredictability are often used to distinguish between systems which are 'complicated' and those which are described as truly 'complex'.

It has been suggested that many health initiatives, mainly because they are dependent on the behaviour of human actors, give rise to what can be described as Complex Adaptive Systems (CAS)[ii] (Ramalingam 2014, (Zhang et al 2014), Paina and Peters 2011, Craig et al. 2008, Rogers 2008, Leykum et al. 2007). Even apparently simple technical interventions can exhibit CAS behaviour as multiple stakeholder groups interact. A CAS will typically exhibit the following characteristics:

- There are a large number of interacting agents.
- Those agents have adaptive capabilities - they can modify their behaviour in ways that impact on the implementation process in response to external influences.
- They will adapt in response to the changing environment - and in particular to changes induced by the intervention and the responses of other agents.
- One common adaptation will be the formation of new alliances that are seen as advancing mutual self-interest.

The implication is that there is no easy way to 'control' or even reliably forecast agent behaviour. Unintended responses to the intervention are common, rendering these systems intrinsically unpredictable. Paina and Peters (2011) argue that the history of attempts at implementing potentially beneficial health systems innovations provides substantial evidence that they should be seen as having these characteristics. Many interventions that were very successful on a small scale, in a research setting, or in one country or region, have often failed when replicated elsewhere or on a larger scale. The implementation of these interventions has rarely proceeded according to plan and in many cases has had to be radically adapted to overcome unforeseen barriers resulting from a rapidly changing environment relating to emerging stakeholder perceptions, attitudes and behaviours. The ability of implementation managers to exercise control over the behaviour of providers, communities and even their own staff has often proved to be highly constrained. Many implementations have displayed a classic characteristic of CAS behaviour, with major inputs sometimes resulting in very limited outcomes or relatively small stimuli having major positive or negative consequences.

CAS Behaviour
As indicated above, CAS can display unexpected behaviours. Three types of behaviour that may be particularly relevant to health interventions are:

1. Feedback loops. An output of a process within a system is fed back as an input into the same system. Example: corrupt behaviour may provide the resources for a health provider to bribe officials, allowing that provider to increase their illicit income and bribe more officials. Example: improving service quality may lead to greater demand for services, which leads to increased facility income, which is used to incentivise staff, which results in improvements in service quality.

2. Path dependence. Processes may have similar starting points and procedures, yet lead to different outcomes, because those outcomes are sensitive not only to initial conditions, but also to the historical context and events that occur over the implementation period. Example: the introduction of rural health insurance works well in one area but not in another where communities remember a similar scheme that was introduced some years ago but failed due to poor management. Example: a brief period of civil unrest in a district when providers are being recruited deters women applicants and results in a health workforce that is overwhelmingly male. This severely constrains the implementation of an innovative reproductive health service for many years as women are reluctant to seek help from a male provider.

3. Emergent behaviour. When two or more agents join together, the resultant alliance may behave in ways which are totally unexpected. Example: health workers who feel they are adversely affected by an intervention come together to form an organisation that can exert pressure on local politicians to delay key aspects of the implementation. Example: community leaders and local health centre managers are made jointly responsible for district drug stores to improve accountability. Many form alliances with traders to sell the drugs in local shops and share the profits. Emergent behaviours are often associated with feedback loops. In the first example, as the number of aggrieved health workers increases so will the number of conversations complaining about the intervention, which will lead to more aggrieved workers. Eventually the system may reach a 'tipping point', when the number of such workers emboldens them to protest formally by forming a new organisation.

4. The butterfly's wing effect. In a CAS, apparently very minor changes in implementation processes can have a substantial effect on outcomes. Example: what the project management team regard as a marginal amendment to an incentive payment scheme is seen as breaking an agreement with providers and is used by a group that opposes the intervention to gain support for withdrawal from the scheme.

> **Example**: In an exploration of the factors influencing immunization coverage in Uganda, Rwashana et al. ([2009](#)) use what they describe as a qualitative systems dynamics approach to illustrate the "*complexity and dynamic nature of the immunization process*" p95. They construct a model of the immunization process, identifying both supply and demand factors and the influences which determine those factors. Figure 4 shows a causal loop diagram illustrating the complexity of the demand side and in particular the multiple and interacting feedback loops that influence the level of demand.

Figure 4: Causal loop diagram - demand for immunization
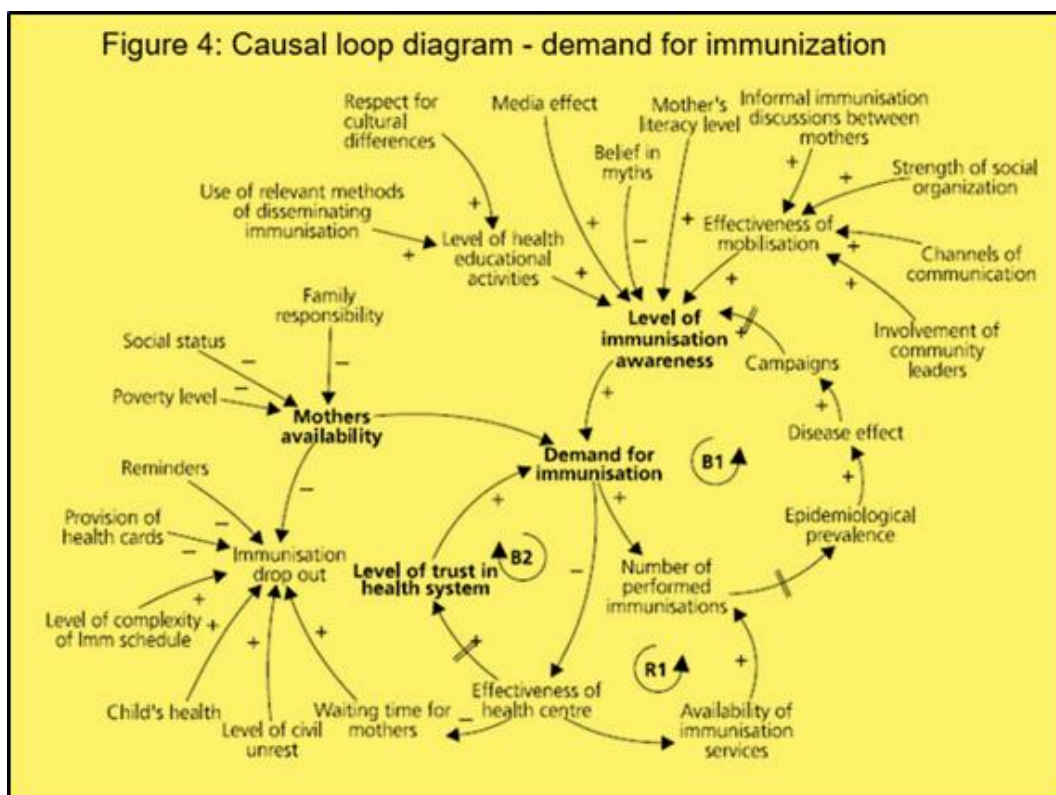
> **Example**: A study of performance-based contracting (PBC) in Uganda ([Ssengooba et al. 2012](#)) argues that previous evaluations focused on the effects of PBC (black-box), paying only limited attention to how these effects arise. Two related theories, complex

adaptive system and expectancy theory were employed. A prospective study tracked the implementation of PBC while collecting experiences of participants at district and hospital levels. It was found that significant problems were encountered in the implementation of PBC that reflected its inadequate design. As problems were encountered, hasty adaptations resulted in a de facto intervention distinct from the one implied at the design stage. For example, inadequate time was allowed for the selection of service targets by the health centres yet they got 'locked-in' to these poor choices. The learning curve and workload among performance auditors weakened the validity of audit results. Above all, financial shortfalls led to delays, short-cuts and uncertainty about the size and payment of bonuses.

Implications of CAS aspects of Health Interventions

From the perspective of the team managing an implementation, the possibility that they will have to address CAS behaviours emphasises the need for flexibility and a willingness to adapt procedures to address unpredicted developments, but only to the extent, as discussed in the opening section of this chapter, that such adaptations do not threaten key elements of the intervention. Defending those elements may require more 'thinking-outside-the-box' in terms of the nature of such threats. For example, rather than simply considering which individual stakeholders are in a position to disrupt or hinder core intervention components, there will be a need to think about the implications of potential alliances forming between two or more stakeholder groups.

As indicated above, feedback loops can be damaging or beneficial to the implementation process. One common mechanism for their emergence relates to informal communications between stakeholders, sometimes based on misunderstandings or exaggerations. By timely and effective communications management, it will often be possible to identify and defuse potentially damaging feedback loops and it may even be possible to encourage those which benefit the implementation. For example, there should be frequent exchanges of information on current and planned activities with affected stakeholder groups, using the communications formats most appropriate to each of those groups. A similar strategy may be applied to instances of emergent behaviour, exploring ways of promoting helpful self-organization and innovation, for example by providing opportunities for increased involvement in implementation management, and at least monitoring the development of alliances which may pose additional threats.

One evident implication of assuming that an intervention may result in CAS behaviours is an even greater need for the effective and timely use of data in planning, adaptation, and evaluation over the lifetime of an implementation. Given the probability that unexpected outcomes will occur, it is essential to establish monitoring systems that can identify when such outcomes start to emerge and track their development over time.

Possible IR concerns when health interventions involve CAS

Given the primary objective of IR as defined in chapter one, the provision of evidence-based insights into the advisability of scaling-up or relocating potentially successful interventions, evidence of CAS behaviour raises a series of additional challenges. There will be a need to assess the underlying determinants and extent of system complexity. For example, does it arise because of the number and heterogeneity of the stakeholder groups involved, from institutional factors relating to the relationships between those groups, or from the previously unrecognised but intrinsic characteristics of the intervention? Which of these are simply unavoidable and which might be mitigated by the design or more effective management of future implementations? There is also an issue as to how to assess the performance of such an intervention. To what extent should unpredicted outcomes, advantageous or detrimental to the intended beneficiaries, be seen as likely to be repeated at scale? Might it be possible to make modifications to the implementation design such that the former were encouraged and the latter discouraged?

The possibility that an intervention may exhibit CAS behaviours emphasises the need discussed in chapter one for long term engagement with the implementation by researchers who work alongside the implementation team. Reconstructing the source and development of such behaviours after an extended interval can be extremely problematic, especially for an external observer. As indicated above, complexity typically arises not from the technical components of an intervention, but from the responses of the various stakeholders and the interactions between them. Typically unexpected developments will give rise to multiple narratives, each shaped by the perceptions, attitudes and motivations of those stakeholders. Only a researcher with in-depth experience of the implementation process and the actors involved might be in apposition to disentangle those narratives and make an informed judgement as to the actual sequence of events and the underlying mechanisms that initiated them.

References

Lucas, Henry, David Evans and Katherine Pasteur (2004). *Research on the current state of PRS monitoring systems*. Discussion Paper 382. Institute Of Development Studies. https://www.ids.ac.uk/files/Dp382.pdf

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science* 2. https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-2-40

Cooper, H., & Geyer, R. (2008). Using 'complexity' for improving educational research in health care. Social Science & Medicine 67:177-182.

Craig, P., Dieppe, P., Macintyre, S., Mitchie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 337:979-983. http://www.bmj.com/content/337/bmj.a1655

DFID (2011). Guidance on using the revised Logical Framework. DFID practice paper. Department for International Development. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253881/using-revised-logical-framework-external.pdf

Gasper, Des. (1997). Logical frameworks: a critical assessment. Managerial theory, pluralistic practice, Working Paper Series No. 264. Institute of Social Studies. http://pdf2.hegoa.efaber.net/entry/content/902/logical_framework_a_critical_ISS.pdf

Grantcraft (2006). *Mapping Change: Using a Theory of Change to Guide Planning and Evaluation*. http://www.grantcraft.org/assets/content/resources/theory_change.pdf

International Network on Strategic Philanthropy (2005). *Theory of Change Tool Manual*. http://www.dochas.ie/Shared/Files/4/Theory_of_Change_Tool_Manual.pdf

Jordon, M., Lanham, H. J., Anderson, R. A., & McDaniel, R. R. J. (2010). Implications of complex adaptive systems theory for interpreting research about health care organizations. Journal of Evaluation in Clinical Practice, 16, 228-231.

Leykum, L. K., Pugh, J., Lawrence, V., Parchman, M., Noel, P. H., & Cornell, J. (2007). Organizational interventions employing principles of complexity science have improved outcomes for patients with Type II diabetes. *Implementation Science* 2(28). http://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-2-28

Mackenzie, Mhairi and Avril Blamey. 2005. The Practice and the Theory: Lessons from the Application of a Theories of Change Approach. *Evaluation* 11(2):151-168. https://www.researchgate.net/publication/249743767_The_Practice_and_the_TheoryLessons_from_the_Application_of_a_Theories_of_Change_Approach

Paina, L., & Peters, D. H. (2011). Understanding pathways for scaling up health services through the lens of complex adaptive systems. Health Policy and Planning 2011:1-9. http://heapol.oxfordjournals.org/content/early/2011/08/05/heapol.czr054.full.pdf+html

Patton, M. Q. (2011). Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. New York and London: The Guildford Press.

Perez, D., Lefevre, P., Castro, M., Sanchez, L., Toledo, E. M., Vanlerberghe, V., et al. (2011). Process-oriented fidelity research assists in evaluation, adjustment and scaling-up of community-based interventions. *Health Policy and Planning* 26:413-422. http://heapol.oxfordjournals.org/content/26/5/413.full

Ramalingham, Ben (2014). *Complex adaptive systems for health systems strengthening.* Complex Adaptive Systems Workshop Presentation. Johns Hopkins University, Baltimore June 2014. www.youtube.com/watch?v=Eqx9DuTAqJM

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation* 14:29-48. http://evi.sagepub.com/content/14/1/29.full.pdf+html

Rwashana, Agnes Semwanga, Willeese Williams Ddembe and Stella Neema (2009). System dynamics approach to immunization healthcare issues in developing countries: A case study of Uganda. *Health Informatics Journal* 15(2):95-107. https://www.researchgate.net/publication/26245581_System_dynamics_approach_to_immu nization_healthcare_issues_in_developing_countries_A_case_study_of_Uganda

Ssengooba, F., McPake, B., & Palmer, N. (2012). Why performance-based contracting failed in Uganda. An 'open-box' evaluation of a complex health system intervention. *Social Science & Medicine* 75:377-383. https://www.researchgate.net/publication/224913346_Why_Performance-Based_Contracting_Failed_in_Uganda-An_'Open_Box'_Evaluation_of_a_Complex_Health_System_Intervention

Vogel, Isobel. 2012. *Review of the use of 'Theory of Change' in international development: Review Report.* UK Department for International Development (DFID).

www.dfid.gov.uk/r4d/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf

W.K. Kellogg Foundation. (2004). *Logic Model Development Guide.* https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide

Zhang, Xiulan, Gerald Bloom, Xiaoxin Xu, Lin Chen, Xiaoyun Liang and Sara J Wolcott (2014). Advancing the application of systems thinking in health: managing rural China health system development in complex and dynamic contexts. *Health Research Policy and Systems* 12:44:1-9. https://health-policy-systems.biomedcentral.com/articles/10.1186/1478-4505-12-44

Chapter 5. Context: Taking situation analysis seriously

Henry Lucas
Institute of Development Studies

1.   Why is context important?

Traditional scientific experimentation typically involves (a) articulation of a plausible theory and (b) testing that theory under carefully controlled conditions to determine if predicted outcomes are observed. We can define 'carefully controlled conditions' as efforts by the experimenter to exclude every factor that could plausibly influence those observed outcomes to an extent that would be of concern, given the objectives of the experiment. For example, in attempting to estimate the acceleration of a falling body due to gravity an experimenter would have to decide whether to conduct the experiment in a vacuum to eliminate the influence of air friction. Their decision would depend on the type of falling body and the required precision of the measurement of its acceleration. Similarly, a chemist intending to measure the heat dissipated in a chemical reaction would have to consider the degree of purity of the chemical compounds involved. Would the measurements be significantly affected if they were only guaranteed to be 99.8 per cent pure as compared to 99.9 per cent? Note that there will be a multitude of other factors - for example the colour of the falling object or the age of the laboratory assistant mixing the reagents - that the experimenter may regard as obviously not relevant to the outcome of the experiment and therefore not needing to be controlled, though as scientific knowledge progresses there is always a possibility that one of these assumptions will later be proved incorrect.

The clinical trials of a new pharmaceutical will also typically involve the use of a range of controls that attempt to isolate the association between application of the drug and observed outcomes, in terms of physiological or psychological changes in a patient, from other factors that might 'confound' that association. As above, the purity of the drug will be carefully assessed. There will be procedures that aim to ensure that patients take their medicine in the prescribed doses, at the appropriate time and in the manner – for example before or after meals – laid down by those organising the trial. Typically, a placebo treatment will often be used to 'control' for the potential effects on patients of simply being involved in a trial, often with neither patients nor providers aware of which patients are receiving the placebo and which the drug until the trial is ended (an approach known as 'double blinding' (Shultz and Grimes 2002)). Somewhat more controversially, the patients will usually be carefully screened before recruitment. They will generally be within a predetermined age range, have no pre-existing relevant health conditions and not be using other medications that may influence the outcome of the trial. They may also be excluded on the basis of a variety of other factors such as their weight, alcohol consumption, smoking habits or other lifestyle behaviours.

Interestingly, many practising physicians have expressed a concern (Zwarenstein and Treweek 2009) that the vast majority of clinical trials can be described as 'explanatory' (designed to test a hypothesis in a highly controlled context), rather than 'pragmatic' (designed to identify treatments that are likely to produce beneficial outcomes across the broad spectrum of patients routinely encountered by healthcare providers). They argue that by adopting 'laboratory' conditions and excluding patients with attributes that might confound the relationship between treatment and outcome, explanatory trials often produce findings that may be of scientific interest but are of limited practical value to clinicians working 'in the real world' and having to make difficult decisions about the best course of treatment for the large number of their patients who do not or will not conform to the rigid guidelines laid down by the drug manufacturers.

2. Context in health systems research

In the health systems interventions with which we are concerned there is almost no possibility of controlling for potentially confounding factors. "In laboratories scientists create artificial conditions in which those causal mechanisms which they conjecture to exist will be activated. In the natural world, potential causal mechanisms will only be activated if the conditions are right for them" (Tilley 2000:5). Interventions take place within a specific context, and implementation successes and failures can often be linked to uncontrolled and often uncontrollable mediating factors that derive from that context (Belaid and Ridde 2015). In a small minority of relatively simple interventions it may be possible to adopt a version of the placebo approach indicated above by randomly allocating individuals in the targeted population to intervention or 'control' groups.[iii] In other cases 'cluster randomisation' may be possible, where whole facilities, villages, health districts, etc. in a targeted region are randomly allocated to receive or not receive the intervention. More often, when random allocation is not seen as a feasible option, the intervention population may be compared with 'similar' populations (with similarity based on the values of a range of socioeconomic and other indicators) that have not received the intervention, in what are usually described as 'quasi-experimental' implementation designs (Gasparrini and Bernal 2015, Harper et al. 2015). In each case the argument (which may or may not be convincing) is that the contexts in the intervention and control groups are sufficiently alike that different outcomes can be attributed to the intervention.

Whatever the intervention design, the implementation team will obviously wish for a successful outcome. To increase the likelihood of achieving this, given their inability to control contextual factors, they should: (a) determine what the most important of those factors are and how and to what extent they might influence outcomes; and (b) find ways to embrace those which are supportive and mitigate those which pose a threat to the implementation process. For the implementation researcher, as discussed in Chapter 1, the tasks are similar but even more challenging. In addition to the above, they would have to: (c) review the extent to which similar factors might need to be addressed in scaling up or relocating the intervention; (d) consider the implications for contexts where some positive factors may be less influential, absent or even negative; and (e) assess the possibilities for using approaches similar to those adopted in the current implementation for the mitigation of negative factors in other contexts.

To take a simple example, those implementing a mother and newborn child health (MNCH) intervention might find that a large majority of their target population have mobile phones (a potential positive factor) but also that a substantial number live in areas where road access is much more limited than expected (potential negative factor). The implementation team might decide to modify their operational procedures to make maximum use of mobile phone communications and to substitute motorcycles for ambulances to overcome the lack of road access. The implementation researcher would also need to consider: the extent to which these factors might be important in other regions; whether implementation performance might be less impressive in locations with more limited communications; and whether motorcycles would be a plausible solution to similar road access limitations in other parts of the country.

It is important to recognise that health system interventions are essentially *social* interventions, and that the diverse range of individuals who make up those societies may respond in very different ways depending both on their specific circumstances and on their perceptions of the intervention. They will often play the most important role in defining the context within which a given implementation takes place. Those contexts will also be strongly influenced by the nature of the communities within which those individuals live. For example, a child health promotion programme may aim to provide information, encourage trust in local services and empower mothers to take healthcare decisions. Programme implementation may trigger different processes depending on the characteristics of targeted individuals and households

(age, education, socioeconomic status, family circumstances, etc.), and various community and societal factors (community assets, local power structures, cultural norms, etc.). Across such varied contexts, the same programme components might in some cases result in increased knowledge and improved attitudes and behaviours, and in others promote conflicts between and/or within households that risk impacting adversely on children's health.

Developing a detailed understanding of the context within which an implementation takes place, and of the actual and potential consequences for implementation progress and outcomes, can thus be seen as one of the defining tasks of the implementation researcher. It may seem a daunting undertaking, given the range of potentially relevant contextual factors that will need to be considered and the limited resources that are typically available. However, remember that the definition of implementation research proposed in Chapter 1 assumes that the researcher will be an active member of the implementation team. This implies: (a) that the work can be shared across a number of individuals, all of whom will (or should) be equally concerned to understand the context within which they are working; and (b) that contextual knowledge can (and should) be acquired over an extended period, not in a 'one-off' exercise.

In practice, the problem faced by both researchers and the implementation team is rarely a lack of available data. As discussed in Chapter 3, even a cursory review of the literature or an elementary internet search will typically uncover a wealth of documentary material relating to the remotest regions and apparently most isolated populations. The difficult task is to identify the often small proportion of that material that provides data that is both relevant and trustworthy. There will also be a large number of individuals - colleagues, professional and social contacts, officials, journalists, etc. - who may be willing to provide key insights into areas that are less well addressed in the literature. For example, an anthropologist colleague assisted one of the editors of this volume by explaining that the design of a project could be relatively easily modified to avoid antagonising a local secret society that might otherwise have persuaded its members to hinder the implementation process. Such informal communications can be invaluable and often obtained with minimal effort - if the researcher has the initiative to seek them out and the ability to assess their reliability.

Whatever the available sources of data, it should be remembered that 'working hard' is no substitute for 'working smart'. It is very easy to lose sight of the primary objective, gathering and interpreting contextual information that is *likely to be relevant to the implementation process*, and to waste valuable time and effort on readily available and interesting, but at best marginally useful, sources. For example, it can be fascinating to investigate the various manuals available in most ministries of health setting out the precise regulations governing the activities of various types of health provider, but if those regulations are routinely disregarded and there is no prospect that they will be monitored or enforced within the lifetime of the implementation, the resources allocated to that investigation should be strictly limited. A useful concept from participatory methodology is that of 'optimal ignorance' - a state achieved when it is recognised that the value of the resources required to gathering additional information will probably exceed the likely benefits (Longhurst 2013).

Sensitive information

The 'secret society' example mentioned above raises an issue that is rarely addressed in textbooks but is often of critical importance - the extent to which potentially sensitive information on contextual factors should be disseminated. It will often be the case that the context within which an implementation is undertaken includes factors that may be acknowledged in private discussions but that would cause serious offence if made a matter of public record. For example, it might become evident that corrupt practices by providers were being tolerated by health authorities, or that some communities were willing to pay for healthcare for male children but not for girls. It would be a matter for the implementation team as a whole to decide how to address such issues. In most cases, a confrontational approach, proclaiming their concerns and endeavouring to overturn long-established practices within a

relatively short time frame, will not be seen as the most effective strategy. Typically, various mechanisms may be introduced into the implementation design, for example modifications to financial control systems or campaigns intended to encourage greater utilisation of services by girl children, and will be described as general project enhancements, without reference to the specific, sensitive problems that they are intended to address.

This situation will often pose a dilemma for the implementation researcher. As a member of the implementation team, it would be entirely inappropriate for them to widely disseminate sensitive information against the wishes of that team. However, given the broader objectives of implementation research as defined in Chapter 1, they clearly cannot ignore evidence that might have serious implications in terms of the potential risks and benefits associated with scaling up the intervention. As will be discussed in Chapter 12, one way to address this dilemma is to move from a focus on 'dissemination', which we commonly associate with academic research findings, to one on 'influencing', which is more relevant to research that is specifically intended to feed into policy decisions. This involves a recognition that the knowledge we have gained from our research is, to adopt a concept from economics, an *intermediate good*, of value only to the extent that we use it to influence policy debates in ways that can be expected to improve health systems and ultimately raise the health status of the population.

From this perspective, the use of sensitive information, just as with any other information, should involve: (a) rigorously determining that you really do have valuable evidence that can and should contribute to policy debates - it is always very tempting to believe that you have unique insights; and (b) presenting that evidence to the relevant audiences in ways that are most likely to influence those debate as intended. Again, direct confrontation will generally not be the most effective strategy in terms of persuading key stakeholders as to the value of your evidence and may well have the opposite effect. Remember that senior officials and politicians will often be well aware of the issues you are addressing and are typically very adept at 'reading between the lines'. In some cases, 'speaking truth to power' may be the best and most courageous option. But you have to be very sure that you are taking this line because it offers the best chance of achieving your ultimate goal and not because it offers the greatest personal satisfaction.

3.   Frameworks for implementation context analysis

As indicated above, context analysis will often need to consider a wide range of factors, some in considerable detail, others simply to confirm that they are likely to be of at most marginal relevance in terms of their influence on the implementation process. In order to undertake a systematic analysis it is of considerable advantage to work within a predetermined framework. Such a framework is best compiled as a collaborative exercise. This should involve at least all members of the implementation team but can often be improved by working with a range of other stakeholders - health officials, providers, community members, etc. - who have specific knowledge of contextual factors that may otherwise be overlooked.

Frameworks will be intervention-specific. For example, a Situation Analysis Tool developed by the Centre for Public Mental Health at the University of Cape Town focuses on the provision and utilisation of mental health services. A guide to situation analysis produced by the Health Systems Trust in South Africa was intended for use by district officials and is therefore primarily concerned with assessing priority health issues and district-level facilities, human resources and management. More recently, the WHO has produced a Situation Analysis Toolkit for the implementation of interventions on male circumcision, which emphasises the need for detailed assessment of local customs and stakeholder attitudes. However, it is possible to consider a number of areas that should usually be at least considered in any such framework. These would include:

1. Politics and history
2. Physical environment
3. Population
4. Health needs and services

Politics and history

Over recent years it has been increasingly recognised in the literature that to understand how a health system functions it is essential to know how it, and the context within which it exists, has evolved over time ([Bloom 2014](#), [Grundy et al. 2014](#)). In the language of complexity theory, it is necessary to acknowledge the importance of 'path dependency' ([Paina and Peters 2012](#)). Where there has been a history of projects that promised much and delivered little, perhaps because of weak local governance structures, it may be very difficult to persuade the population that a new intervention will be successful. Where corruption has become endemic, some stakeholders will view such an intervention as a potential new source of funds, while others will be very reluctant to participate, assuming that the benefits will be misappropriated. Where there are long-standing ideological differences between different sections of the population, there will be a risk that any new development will become a cause of dissent between different political factions. On the other hand, in populations that have a history of effective community organisations, such as proactive local women's groups, it may be much easier to set up, for example, a community-based health insurance scheme ([Asaki and Hayes 2011](#)).

As indicated by the above, issues relating to local and national politics and to historical trends and events may well be seen as extremely sensitive and difficult to address within an implementation research setting. However, in many cases they will be among the most important contextual factors. Projects and programmes have come to an abrupt halt when a change of government has removed key political supporters. Others have failed to increase service utilisation because local health officials had lost the trust of a substantial section of the targeted population as a consequence of previous activities. The controversy surrounding the clinical trial in 1996 of a new antibiotic by the drug company Pfizer in Kano, Northern Nigeria, during an epidemic of meningococcal meningitis ([Wise 2001](#)), is still raised by Nigerians as an example of the risks of engaging with foreign companies in the health sector and has played a part in the resistance to polio vaccination programmes ([Yahya 2006](#)). As discussed above, under the heading 'sensitive information', the argument here is not that the implementation researcher should provoke controversy by reopening old wounds or taking sides in any political debate. However, if past events and current political positions are relevant to the potential outcome of the implementation process, they do have to be addressed, analysed and interpreted as an important component of the research findings. Again, the pragmatic use of those findings to influence policy will be discussed in Chapter 12.

Physical environment

The physical environment within which an intervention takes place should almost always play a major role in determining implementation design. To give an extreme example, implementation of an intervention designed to improve health outcomes for children living in the crowded squatter settlements of Nairobi will clearly pose very different problems from one intended for scattered populations in the highlands of Papua New Guinea or the densely populated islands of the Sundarbans mangrove forest in West Bengal. Even when the targeted regions are relatively limited in size, substantial geographical variations within regions may need to be carefully considered. For example, researchers will typically distinguish between urban and rural locations but not differentiate peri-urban areas, which often have their own very specific environmental characteristics. Similarly, it will often be essential to classify rural areas into those that are easily accessible and those that are more remote from major centres

of population, given that it will typically be substantially more difficult to provide services in the latter.

Note that it should be standard practice to explain *why* different environmental factors are relevant. Researchers often provide descriptions that would be more appropriate for a geography textbook or a travel guide, specifying items such as precise estimates of land area, height above sea level, average annual rainfall or detailed grid references. Often the key issues concern potential physical access barriers – such as long, difficult and/or dangerous journeys to services by those seeking care (Houben et al. 2012) or restrictions on the ability of providers to transport medical supplies or appoint additional staff to facilities when required (Cohen et al. 2010). Another important question that is often overlooked relates to the willingness of providers (and their families) to live and work in 'difficult' areas, whether urban shanty towns or remote rural areas (Agyei-Baffour et al. 2011, Sundararaman and Gupta 2011).

Consideration of such issues naturally leads to questions relating to infrastructure and services. We often use some indicator of the overall 'level of development' of an area, such as GDP, but where possible such measures should be supplemented by data on specific factors that are considered relevant to the planned intervention. Is the area well served by road, rail or water transport links? Is there access to electricity, clean water and sanitation? Are there local primary/secondary schools? Is there is a reasonably effective and trusted law enforcement service? Are there functioning telecommunications networks that could enable access to services via telephone, radio or television? How are these various services affected by seasonal factors such as rainfall, drought, high winds or snow?

Population

Just as it is important to distinguish between geographic areas with different characteristics, it is equally important to consider the extent to which distinctions between different population groups will be relevant to the outcome of the intervention. Sometimes there will be a considerable overlap between geographical areas and population characteristics. For example, in Nigeria certain states in the south-eastern region are closely associated with the Igbo ethnic group, the majority of whom identify with the Christian religion. Even in such cases the researcher should be very cautious in assuming that the number of individuals in that area who do not share those characteristics is insignificant. The 2013 Demographic and Health Survey for Nigeria suggests that almost 98 per cent of the population of the south-eastern region are Igbo. That would still imply, however, that around half a million individuals who identify themselves as being in other ethnic groups also live in that region. Given that resources are always constrained, implementers may sometimes reasonably decide that they will tailor an implementation process in ways that they see as most likely to meet with approval from the majority population in a given location, even though this may adversely affect the response from minority groups. Nevertheless, any such decision should be clearly stated, justified and evidence-based.

Often it will be evident that there are relevant differences between population groups living in the same geographical area. For example, as in many other countries, most major cities in China have large migrant populations. Those populations have very limited access to the services, including health services, provided for those who have urban resident status (Mou et al. 2013). Any intervention intending to improve the health of the overall population living in a city would have to consider the very different circumstances of these two groups when developing the implementation design. Similar considerations would apply in south-western Nigeria, where there are many long-established settlements entirely composed of members of the Hausa ethnic group in cities that are predominantly populated by the Yoruba people. As the purpose of these settlements was precisely to retain traditional customs and practices, including authority structures, the context within which they live differs substantially from that of the majority population (Omobuwa et al. 2013).

Having identified relevant population groups on the basis of factors such as geographical location, migrant/non-migrant, ethnicity, culture, religion, etc., it will also be essential to consider the extent of variation within these groups. Gender will almost always be a key factor. Where an intervention is intended to improve the health of all members of a population group, for example advocating behavioural change to reduce the risk of chronic illness, it seems obvious that careful thought should be given as to how such an intervention will be perceived by both women and men and how they will respond. However, this should be the case even if the intervention is clearly gender-specific, for example intended to encourage increased use of ante-natal care, given that such perceptions and responses are invariably strongly influenced, positively or negatively, by existing gender relations (e.g. Dworkin et al. 2012, Nikiema et al. 2012). Potential differences between the younger and older members of population groups should also be considered. These may include attitudes to cultural traditions, authority structures or technical innovations. For example, a number of studies have suggested that older people are much less likely to use mobile phone texting services (e.g. Deng et al. 2014), with implications for behaviour change interventions that wish to adopt this approach. Note that it will often be informative to consider age and gender simultaneously. An intervention to encourage facility-based births, for example, might face opposition from older women who trust local traditional birth attendants with whom they have shared life experiences.

Variations in levels of education may also be an important consideration. Some interventions, for example those that provide written instructions and/or warnings to patients in relation to drug treatments, may be premised on the assumption that the great majority of the targeted population are either literate in one of the languages selected for use in the intervention or will be able to rely on support from a literate person. If this is not the case, alternative approaches may have to be adopted (Dowse and Ehlers 2001). There is evidence that adherence to anti-retroviral therapy can be influenced by the level of education of patients, though not always in the expected direction (Emamzadeh-Fard et al. 2012, Radhakrishnan et al. 2012). Interventions that involve some form of written contractual relationship – for example where individuals are invited to join a health insurance scheme – may be more easily comprehended and thus more attractive to those with language and/or numeracy skills beyond those that would be acquired at the primary level of education (Jehu-Appiah et al. 2011). A general understanding of the distribution of the population across different education levels may therefore be useful in predicting the potentially different responses to various components of an intervention, with implications for implementation design.

Finally in this section, we need to consider potential financial barriers to care. Even when an intervention is providing a notionally 'free' service we will typically find that utilisation is significantly lower for the poorer members of the population. This may sometimes be because facilities find ways to add indirect fees, for example to register with the facility, or because they encourage patients to take additional services - laboratory tests, scans, supplementary drugs, etc. - and imply that this will greatly increase the efficacy of the basic, free treatment. In some cases, it may be that patients have to bear travel or accommodation costs, or, more recently, need access to a mobile phone to utilise the service. It is also possible for the 'opportunity costs', associated with a household member having to take time away from wage employment, household production or other tasks, to be *relatively* high for poor households. In poor areas of rural China, for example, poverty is often associated with a lack of household labour time because household sizes tend to be limited and many of those of working age leave to seek employment in urban areas.

The aim of contextual analysis in this area would be to gain an understanding as to which sections of the identified population groups might either fail to access treatment due to financial barriers, or experience serious hardship due to expenditures associated with accessing treatment - often described as 'catastrophic healthcare expenditure' (Mills et al.

2012). This will typically involve compiling order-of-magnitude per capita annual household income or expenditure estimates, focusing on those within the identified groups most at risk. These might include, for example, small farmers and the landless in rural areas, day labourers and the self-employed in urban areas. These estimates will sometimes be available from sample survey data but it may often be necessary to rely on the judgement of a number of key informants. For interventions where substantial out-of-pocket expenditures might be involved - for example a co-payment scheme for inpatient treatment - it will also be useful to explore the extent to which these groups tend to have disposable assets, outstanding debts and access to sources of credit (including from extended families). In many countries, illnesses that result in substantial costs or loss of income can severely disrupt the livelihoods of households that have to sacrifice productive assets, including agricultural land, or take out loans on highly unfavourable terms that may force asset sales at a later date (de Laiglesia 2011, Kenjiro 2005). Finally, there should be at least some consideration of intra-household financial arrangements. For example, in many countries expenditure on healthcare for children may be influenced by women's status in decision-making and control over household resources (Richards et al. 2013).

Health needs

An intervention intended to improve the detection and treatment of TB cases should obviously compile as much information as possible about the incidence of TB in the target population, the extent to which those with TB have access to services and the extent to which they utilise those services. This would apply to any disease-specific intervention. Even for such focused interventions, however, it will often be important to compile data on a range of other health issues. Such data can, for example, assist us to understand population awareness and perceptions of the health issues with which we are primarily concerned, which may help explain their attitudes to our intervention. For example, a population in which both adults and children are subject to frequent bouts of fever, cough and diarrhoea may question why those implementing an intervention on chronic conditions such as hypertension or diabetes are failing to address what they, and possibly local health service providers, see as more immediate concerns.

More wide-ranging interventions, for example the introduction of performance-related payments in primary healthcare centres or new insurance schemes to meet the cost of inpatient care, would merit the compilation of detailed information on a range of relevant conditions with high incidence or prevalence rates in the target population. Such data should allow an improved understanding both of the healthcare needs of the population and of the current and potential demands placed on healthcare providers. Thus in the example above, reliable estimates of the likely rates of hypertension, diabetes and other chronic diseases in a population will not only indicate the need for an intervention to address such conditions but also the implications for healthcare services of improvements in their detection and diagnosis that may result from such an intervention. Note that cultural factors may complicate the translation of health needs into demands for health services. An obvious example in many countries relates to mental illnesses, where the stigma attaching to mental illness and the assumption that it should be addressed by religious or traditional healers will often prevent sufferers and their families from seeking care from allopathic providers (Brenman et al. 2014).

In many countries reliable data on incidence/prevalence rates and service utilisation over a recent period will be difficult to obtain, except possibly where there are well-funded major programmes for specific diseases such as HIV/AIDS or TB, or where it is possible to obtain access to detailed facility or health insurance data. In general, researchers will have to rely on evidence from previous national surveys. For example, data on the most common early childhood diseases can be found from the Demographic and Health Surveys undertaken in many countries. It should be noted that such surveys are typically based on reported symptoms rather than formal diagnosis and that they rely heavily on the ability of respondents to provide details as to the type of healthcare accessed. A useful international source of data

on disease-specific morbidity and mortality for most countries is provided by the Global Burden of Disease studies. The country profiles compiled under this programme represent systematic attempts to use whatever data are available to derive best estimates of the impact of different diseases based on the number of years of life lost to premature deaths and the number of years lived with a disability. Note that national surveys will aim to provide data for the population as a whole. Disaggregation by location or population group may be possible to some extent, depending on the nature of the survey and whether access to the raw data is possible, but it will often be impossible to derive disease patterns for the specific population targeted by the intervention.

Health systems

In Chapter 1 it was argued that the most useful approach to the study of health systems, especially in situations where there are multiple providers and limited regulation of services (Bloom and Standing 2008), is to require the implementation researcher to define the health system - in terms of agents, units and institutions - that will be the focus of their research. However, in order to undertake that definition, to define the boundaries of the system with which they are concerned, they should first undertake a systematic assessment of the diverse range of providers that are offering to provide health services to the population targeted by the implementation. This may not be a simple task. For example, we often characterise health service providers under headings such as:

- Public healthcare;
- Formal private healthcare;
- Informal private providers (unlicenced practitioners, shops, drug sellers);
- Traditional, religious and faith healers;
- Household healthcare.

However, many of these categories consist of multiple components that have their own distinct characteristics. In China, for example, public hospitals may offer either allopathic (Western) medicine, Traditional Chinese Medicine, or both. Ayurvedic practitioners have recently gained a similar status in India. In the United Kingdom, there are long-standing debates as to whether homeopathic treatments should be available under the public National Health Service. In most countries the 'formal' private sector will include individuals from a variety of medical traditions ranging from senior specialists, with qualifications and experience equal to or greater than those in the public sector, to providers with minimal training and titles ranging from Registered Medical Practitioner in India (Das and Hammer 2007) to Village Doctor in Bangladesh (Mahmood et al. 2014) to Community Health Worker in other countries. In terms of the quality of services provided, there may be little to choose between many of these providers and those who practise without any form of licence or simply sell drugs in shops or local markets. In both groups there will be dedicated, principled providers who sincerely believe that they are doing their utmost to help those who seek their services and there will be unprincipled charlatans or 'quacks', whose primary aim is to extract money from or exert influence over their patients.

Having identified those components that seem most relevant to the implementation - that is, those which play a substantial role in delivering the services on which the implementation is focused - it will be helpful to undertake a systematic descriptive analysis to identify for each component the various units, decision-making agents and (formal and informal) institutions that govern its operation. A simple outline example is provided below (Tables 5.1 and 5.2), where we adopt the WHO classifications discussed in Chapter 1 to consider different aspects of each type of service provider. As always in health systems research, we will wish to assess the implications of our analysis in terms of access, utilisation and quality of services for different population groups. In general, it will be relatively straightforward to identify the main agents and units but understanding the institutional arrangements will typically be more challenging.

A reasonable understanding of the formal institutions can usually be gained by a review of the available documentation: policy statements, plans, laws, regulations, protocols, procedures, guidelines, etc., though identifying and prioritising such documents will often require guidance from key informants. However, in many cases these documents will describe a system that differs substantially from that which actually exists. Undertaking research in Nigeria in the mid-1990s, it was common to find in some local government areas that the rural primary healthcare system described by the ministry of health simply did not exist. Health workers had not been paid for many months and had moved away to seek other employment, equipment had become unserviceable or disappeared and in some cases buildings had collapsed because of long-term failures in basic maintenance. Similarly, following the economic reforms in China, the rural 'three-tier healthcare system', under which County Hospitals supervised Township Healthcare Centres, which supervised Village Health Stations, evolved into what were essentially local competitive markets, with each facility competing for patient 'out-of-pocket' fees. It took many years before this situation was officially recognised and action taken to offset the worst characteristics of these markets. In one East African country, basic drugs were at one time primarily available via 'essential medicine kits' supplied by an international aid agency. It was general knowledge that many of these kits were diverted to local shops to which patients would be referred even by providers at public health centres. However, formal acknowledgement of this practice would have resulted, under the agreement with the agency, in the withdrawal of the kits until more secure delivery mechanisms could be devised. As this was seen as creating a potentially life-threatening situation for many children, everyone proceeded as if they were unaware of the true situation.

In such situations it will be necessary to understand both the intended and the actual institutional arrangements that are governing the activities of a health system. While it may be common knowledge that that system is functioning quite differently from what was intended in the various policy documents and formal operating guidelines, it is rarely possible for those charged with managing the system to radically shift their activities to allow for that fact. Officials, managers, administrators and providers have contracts of employment that assume intended operating procedures. Data collection systems and reporting forms will have been designed to align with those procedures. In some countries, health officials will take great care in preparing annual budgets that they know will be ignored. In many others those managing health information databases will use sophisticated techniques to analyse and present data that are well understood to be incomplete and highly unreliable. Institutions are important even when they result in unintended consequences. The implementation research has to understand both how they were intended to function and how they fail.

Table 1: Public health sector

	Units	Agents	Institutions
Service delivery	Hospitals Clinics Laboratories	Doctors Nurses Technicians Managers	Legal system Professional guidelines Peer review
Human resources	Hospitals Clinics Laboratories Training institutions	Managers Doctors Nurses Technicians Teachers	Legal system Professional guidelines Curricula
Supplies and equipment	Facilities Pharmacies Drugstores	Managers Pharmacists Drug company representatives	Legal system Equipment norms Procurement guidelines Drugstore procedures Audit

Leadership and governance	Ministries Hospital Boards Professional organisations	Ministry officials Hospital directors	Legal system Professional guidelines
Financing	Patients Insurance agencies Government	Patients Insurance agents Ministry officials	Legal system Insurance schemes
Information systems	Registers District reports Health surveys	Administrators Enumerators Statisticians	Data quality audit Dissemination procedures Notifiable diseases

Table 2: Informal private health sector

	Units	**Agents**	**Institutions**
Service delivery	Clinics Shops Markets	Village doctors Drug sellers Shop salespersons	Legal system Custom and practice
Human resources	Households	Village doctors Drug sellers	Custom and practice
Supplies and equipment	Shops Drugstores	Shopkeepers Wholesalers	Legal system Custom and practice
Leadership and governance	Providers Trade organisations	Village doctors	Legal system Custom and practice
Finance	Patients Community organisations	Patients	Markets Custom and practice
Information systems	Account books Receipts	Shopkeepers Drug sellers Purchasers	Custom and practice

4.   Stakeholder analysis

Stakeholder analysis (Brugha and Varvasovszky 2000) can be one of the most important activities undertaken by researchers in terms of understanding the context within which the implementation takes place - but only if it is systematic and comprehensive. As discussed in Chapter 1, even apparently simple technical health system innovations typically involve complex social interventions. The extent to which different groups respond with enthusiasm, indifference or hostility to an implementation will often determine its relative success or failure. On occasion, depending on their degree of authority or influence, a single individual can make the difference. Predicting likely responses, determining their potential implications, adapting procedures in line with that analysis and *repeating this process over the lifetime of the implementation* is a key activity for the implementation team in terms of maximising the likelihood of achieving targeted outcomes. For the implementation researcher, engaging in a rigorous stakeholder analysis can provide valuable insights into the social contextual factors that might promote or obstruct scaling up or relocation of the studied intervention. Too often, however, such analyses are not allocated the resources that they merit, being seen as simply a routine task to be undertaken at the start of the implementation, sometimes simply to meet the requirements of those funding the intervention, and largely disregarded thereafter.

We can outline the aims and objectives of stakeholder analysis as follows (Varvasovszky and Brugha 2000):

> ***Aims: Identify all relevant stakeholders and assess:***
> - how they are likely to be affected by the intervention;

- how they are likely to respond;
- the implications of their responses, given their capacity to influence implementation outcomes either directly or through their relationships with other stakeholders.

Objectives:
- Where possible, modify the implementation design to (a) encourage collaboration and (b) minimise obstruction by different stakeholders;
- Improve understanding of the underlying causes of implementation successes or failures that are linked to stakeholder behaviour.

A stakeholder can be any individual, group or organisation that may be positively or negatively affected by an intervention or in a position to influence the implementation of that intervention and have a positive or negative effect on intended outcomes. Broadly speaking, the analysis is intended to generate information about these stakeholders that can improve our understanding of their incentives, perceptions, attitudes, and relationships with other stakeholders in order to provide insights into their current and likely future patterns of behaviour in relation to the implementation of a given intervention (Hyder et al. 2010). This requires:

1. Identification of potential stakeholders.
2. For each potential stakeholder, determination of:
   a. The extent to which they are interested in the intervention;
   b. How and to what extent they can influence the implementation progress;
   c. Their attitudes and actions relating to the intervention objectives;
   d. The factors that are most important in determining those attitudes and actions.

Identifying stakeholders
One way to identify stakeholders is to work 'outward' from the implementation:

- Start with the project managers or lead implementers and donors/funders;
- Move out to consider those they work with - partners, service providers, regulators and owners of resources (facilities, land, infrastructure, etc.);
- Move out again to those who will either be involved in the activities or who are beneficiaries - NGOs, local authorities, communities, media organisations and other interest groups.

Figure 1: Identifying stakeholders
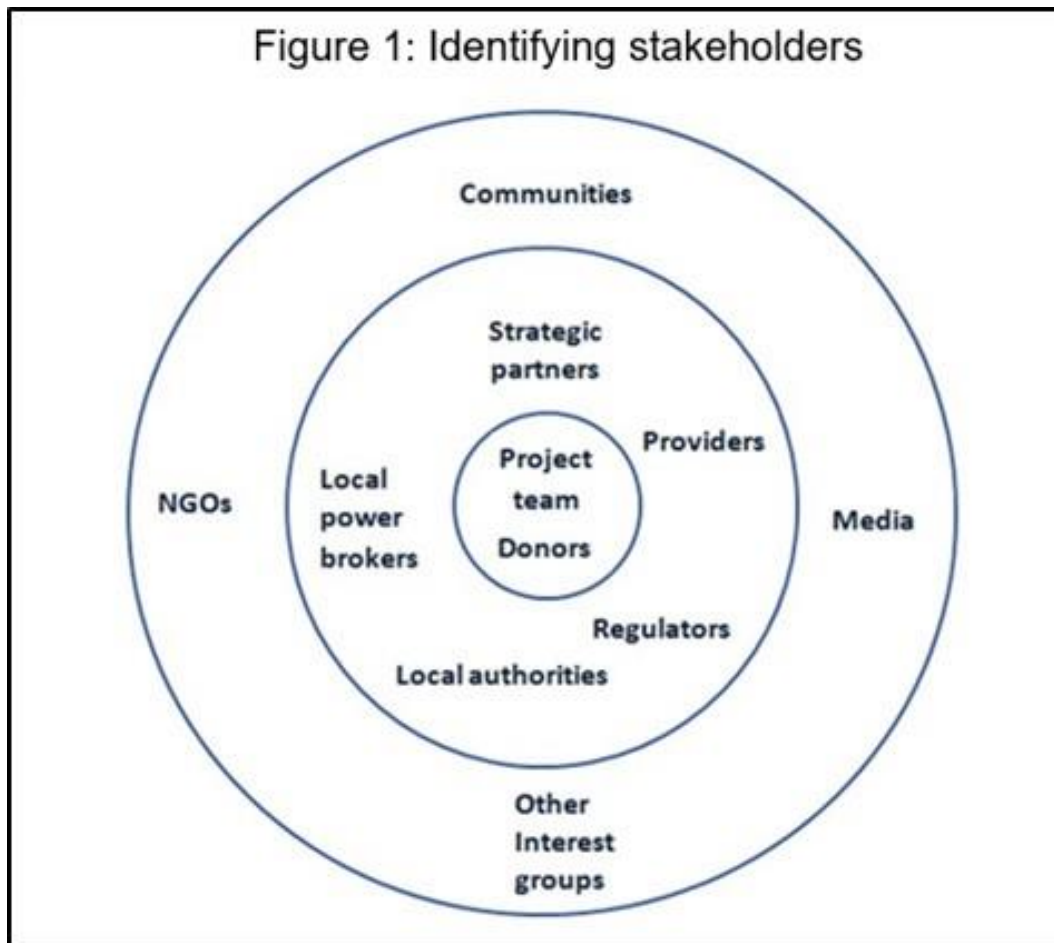
Note that this diagram is intended to assist in the identification of ALL stakeholders - those who may impact ON or be impacted BY an intervention.

Stakeholder groups and subgroups

The limited resources typically allocated to stakeholder analysis often results in a failure to achieve an appropriate level of disaggregation. It is important to distinguish significant stakeholder subgroups that may have very different attitudes and levels of influence. For example:

- Providers should at least be subdivided into categories such as: public/private, qualified/unqualified, allopathic/traditional/faith-based but some degree of cross-classification (e.g. qualified traditional private providers) may substantially increase the value of the analysis by reducing within-group variation.
- Community members might be classified in terms of: male/female, younger/older, richer/poorer, indigenous/migrant, ethnic group, etc., again keeping open the potential value of cross-classification (e.g. older poorer women)
- The large number of potentially interesting groups and subgroups requires a careful process of prioritisation in terms of analysis. The criterion for prioritisation should always be in terms of the anticipated level of influence over implementation outcomes that different subgroups may possess.

Analysis

The analysis will focus on identifying the following characteristics of each stakeholder or stakeholder group in line with the aims indicated above:

1. Their interest in the intervention:
    *   Intended beneficiary;
    *   Direct involvement in the implementation;
    *   Likely to be directly affected by the intervention (positively or negatively);
    *   Likely to be influenced by those directly affected;
    *   No apparent interest (can be omitted from the current list of stakeholders but may have to be added later if this assessment changes).
2. Their potential influence over the implementation process, for instance:
    *   Policymaker (capacity to affect implementation strategy or context);
    *   Decision-maker (capacity to affect routine implementation activities);
    *   Gatekeeper (capacity to control access to resources or take-up of services);
    *   Opinion leader (capacity to affect responses of other stakeholders).
3. Their attitude to the intervention (evidence to be provided where available):
    *   Enthusiastic supporter;
    *   Generally supportive;
    *   Indifferent;
    *   Opposed;
    *   Strongly opposed.
4. Factors driving attitudes, possibility of changing attitudes and potential benefits:
    *   How likely are they to use their influence to change outcomes and why?
    *   What would be the consequences?
    *   Are they accessible to the implementation team?
    *   Might they be responsive to incentives intended to modify their attitudes?

Given the complexities of stakeholder analysis it is sometimes tempting to make the simplifying assumption that a given stakeholder or stakeholder group can be considered in isolation, and that they make their decisions in line with their own perceptions and preferences. However, in the real world we know that this is rarely the case and that relationships between stakeholders, especially power relationships (Erasmus and Gilson 2008), play a major role in influencing behaviour. A hospital manager may be convinced that a new payment mechanism would result in a better outcome for patients but act to undermine that mechanism because he wishes to avoid conflicts with senior hospital staff who believe that it will adversely affect their incomes. Local government health officials may see the training, licensing and monitoring of local drug sellers as the most effective way to deliver anti-malarials but be pressured into opposing this reform by qualified providers with connections to local politicians, who see it as threatening their control over the supply of prescription drugs.

This indicates the need to identify stakeholder networks (Blanchet and James 2012). The aim will be to map the formal and informal links between stakeholders and assess the underlying nature of those links - in particular do they exist only in theory (e.g. according to regulations) or do they have practical consequences? Relationships may be of many types including: financial support; direct management; oversight/monitoring (in theory and in practice); advice/influence (in theory and in practice), etc. Having identified such relationships, the stakeholder analysis can be revisited to explore the extent to which they provide additional clues to the attitudes and behaviours of particular stakeholders.

Data collection

A range of data collection activities will need to be undertaken. The aim will be to seek: responses to a range of specific questions; the reasons underlying these responses; and the extent to which the responses are based on available evidence. The initial activity should involve a detailed document review to assess the stated position (if any) of each stakeholder on issues relevant to the intervention objectives. This will be followed by primary data collection using semi-structured interviews, structured questionnaires (possibly self-administered) and focus groups.

One important consideration is the extent of involvement in data collection, feedback and analysis by the various stakeholders themselves. Such involvement can offer many advantages in terms of the extensive knowledge that individual stakeholders may have, for example in terms of the range of attitudes inside an organisation or internal documents that may be difficult to identify by other means. On the other hand, there are obvious risks that some stakeholders may attempt to drive the analysis in directions that support their own agendas. Decisions should be made on a case-by-case basis depending partly on the sensitivity of the information compiled. Where there are no concerns that the analysis may cause offence, a feedback stage can be included in the process to allow each stakeholder to comment on findings relating to their own position and correct factual inaccuracies where appropriate.

Finally, we would suggest that in gathering the data to perform a stakeholder analysis it is useful to keep in mind the work of pioneering sociologist Erving Goffman (Goffman 1956). This codified the commonplace observation that individuals can be compared to actors who play a variety of roles. They will behave very differently depending on whether they are 'front-of-stage', 'offstage' or 'backstage'. Front-of-stage is where the actor formally performs and here they will adhere to conventions that align with the expectations of their current audience. For example, a hospital director accompanied by senior staff members responding to questions from an unfamiliar researcher would probably play a very different role from that which they would adopt if called to a formal meeting with the Minister of Health. Offstage is where actors meet individual audience members in an informal environment. Thus, if our hospital director happened to meet the researcher at a social gathering and realised that they were both long-standing friends with the local politician who was the host of that gathering, he or she might well be much more open in discussion of current problems in the health sector. Even so, this would be another role that the director was playing. Only when they (and we) are alone 'backstage' do actors truly get to be themselves.

The lesson to be learned is that it is always wise to be cautious about the extent to which any researcher can truly 'understand' the perceptions and attitudes of different stakeholders. Those who proclaim themselves to be strongly in favour of an intervention may act in ways that undermine its implementation and those who are initially most critical may become key players in ensuring its success. Only by updating the stakeholder analysis on a regular basis and comparing actions to stated intentions can the implementation researcher hope to gain at least a working understanding of the factors influencing stakeholder behaviour.

References

Agyei-Baffour, P., Kotha, S.R., Johnson, J.C., Gyakobo, M.; Asabir, K., Kwansah, J. et al. (2011). Willingness to Work in Rural Areas and the Role of Intrinsic Versus Extrinsic Professional Motivations: A Survey of Medical Students in Ghana. *BMC Medical Education* 11:56. http://bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-11-56

Asaki, B. and Shannon, H. (2011). Leaders, Not Clients: Grassroots Women's Groups Transforming Social Protection. *Gender and Development* 19.2: 241-53. http://www.tandfonline.com/doi/abs/10.1080/13552074.2011.592634

Belaid, Loubna and Valery Ridde (2015). Contextual factors as a key to understanding the heterogeneity of effects of a maternal health policy in Burkina Faso? *Health Policy and Planning* 30:309-321. http://heapol.oxfordjournals.org/content/early/2014/03/14/heapol.czu012

Blanchet, K. and James, P. (2012). How to Do (or Not to Do)... A Social Network Analysis in Health Systems Research. *Health Policy and Planning* 27(5): 438-46. http://heapol.oxfordjournals.org/content/27/5/438.full.pdf+html (accessed 30 January 2015)

Bloom, G. (2014). History, Complexity and Health Systems Research, *Social Science and Medicine* 117: 160   61. https://www.researchgate.net/publication/264248823_History_complexity_and_health_systems_research

Bloom, G. and Standing, H. (2008). Future Health Systems: Why future? Why now? *Social Science and Medicine* 66: 2067   75. http://r4d.dfid.gov.uk/Output/177233/

Brenman, N.F., Luitel, N.P., Mall, S. and Jordans, M.J.D. (2014). Demand and Access to Mental Health Services: A Qualitative Formative Study in Nepal. *BMC International Health and Human Rights* 14(22)1-12. www.biomedcentral.com/content/pdf/1472-698X-14-22.pdf (accessed 30 January 2015)

Brugha, R. and Varvasovszky, Z. (2000). Stakeholder Analysis: A Review. *Health Policy and Planning* 15(3):239   46. http://heapol.oxfordjournals.org/content/15/3/239.full.pdf

Cohen, J., Sabot, O., Sabot, K., Gordon, M., Gross, I., Bishop, D. et al. (2010). A Pharmacy Too Far? Equity and Spatial Distribution of Outcomes in the Delivery of Subsidized Artemisinin-Based Combination Therapies through Private Drug Shops. *BMC Health Services Research* 10: S6. http://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-10-S1-S6

Das, J. and Hammer, J. (2007) 'Location, Location, Location: Residence, Wealth, and the Quality of Medical Care in Delhi, India', *Health Affairs* 26.3: w338-w351, http://content.healthaffairs.org/content/26/3/w338.full.pdf+html (accessed 30 January 2015)

Deng, X., Ye, L., Wang, W. and Zhu, T. (2014). A Cross-Sectional Study to Assess the Feasibility of a Short Message Service to Improve Adherence of Outpatients Undergoing Sedation Gastrointestinal Endoscopy in the People's Republic of China', *Patient Preference and Adherence* 8: 1293-97.

Dowse, R. and Ehlers, M.S. (2001). The Evaluation of Pharmaceutical Pictograms in a Low-literacy South African Population. *Patient Education and Counselling* 45:87   99. http://www.pec-journal.com/article/S0738-3991(00)00197-X/abstract

Dworkin, S.L. Colvin, C. Hatcher, A. and Peacock, D. (2012). Men's Perceptions of Women's Rights and Changing Gender Relations in South Africa: Lessons for Working with Men and Boys in HIV and Antiviolence Programs. *Gender and Society* 26: 97. http://gas.sagepub.com/content/26/1/97.abstract

Emamzadeh-Fard, S., Sahar, E. Fard, S. A., Alinaghi, S. and Paydary, K. (2012). Adherence to Anti-Retroviral Therapy and Its Determinants in HIV/AIDS Patients: A Review. *Infectious Disorders - Drug Targets* 12: 346   56. http://www.ingentaconnect.com/content/ben/iddt/2012/00000012/00000005/art00004?token=00551b2829e8dc67749437a63736a6f5e475f5d7d666a707b3a446f644a467b4d616d3f4e4b34e003ac59

Erasmus, E. and Gilson L. (2008). How to Start Thinking about Investigating Power in the Organizational Settings of Policy Implementation. *Health Policy and Planning* 23: 361-68. http://heapol.oxfordjournals.org/content/23/5/361.full.pdf+html

Gasparrini, A and J L Bernal (2015). Commentary: On the use of quasi-experimental designs in public health evaluation. International Journal of Epidemiology 44(3): 966-968. http://ije.oxfordjournals.org/content/44/3/966.long

Goffman, E. (1959). *The Presentation of Self in Everyday Life.* New York: Doubleday. https://monoskop.org/images/1/19/Goffman_Erving_The_Presentation_of_Self_in_Everyday_Life.pdf

Grundy, J., Hoban, E., Allender, S. and Annear, P. (2014). The Intersection of Political History and Health Policy in Asia: The Historical Foundations for Health Policy Analysis. *Social Science and Medicine* 117: 150-59. http://www.sciencedirect.com/science/article/pii/S0277953614004766

Harper, S, Charters TJ, Strumpf EC, Galea S, Nandi A. (2015). Economic downturns and suicide mortality in the United States 1980-2010: observational study. International Journal of Epidemiology. 44 (3): 956-966. http://ije.oxfordjournals.org/content/44/3/956.long

Houben, R., Van Boeckel, T., Mwinuka, V., Mzumara, P., Branson, K., Linard, C. et al. (2012). Monitoring the Impact of Decentralised Chronic Care Services on Patient Travel Time in Rural Africa   Methods and Results in Northern Malawi. *International Journal of Health Geographics* 11:49. https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-11-49

Hyder, A., Syed, S., Puvanachandra, P., Bloom, G., Sundaram, S., Mahmood, S. et al. (2010). Stakeholder Analysis for Health Research: Case Studies from Low- and Middle-income Countries. *Public Health* 124: 159   66. https://www.researchgate.net/publication/41942614_Stakeholder_analysis_for_health_research_Case_studies_from_low-_and_middle-income_countries

Jehu-Appiah, C., Aryeetey, G., Spaan, E., de Hoop, T., Agyepong, I. and Baltussen, R. (2011). Equity Aspects of the National Health Insurance Scheme in Ghana: Who is Enrolling, Who is Not and Why? *Social Science and Medicine* 72: 157   65. https://www.researchgate.net/publication/49671117_Equity_Aspects_of_the_National_Health_Insurance_Scheme_in_Ghana_Who_Is_Enrolling_Who_Is_Not_and_Why

Kenjiro, Y. (2005). Why Illness Causes More Serious Economic Damage than Crop Failure in Rural Cambodia. *Development and Change* 36.4: 759-83. https://www.researchgate.net/publication/229575456_Why_Illness_Causes_More_Serious_Economic_Damage_than_Crop_Failure_in_Rural_Cambodia

de Laiglesia, J.R. (2011). Why Do Farmers Sell their Land? Evidence from Nicaragua. OECD Development Centre. de Laiglesia, J.R. (2011). Why Do Farmers Sell their Land? Evidence from Nicaragua. OECD Development Centre. pagesperso.dial.prd.fr/dial_pagesperso/dial.../204_de%20laiglesia.pdf (accessed 30 January 2015)

Longhurst, R. (2013) 'Implementing Development Evaluations under Severe Resource Constraints', Centre for Development Impact Practice Paper 3, Brighton: Institute of Development Studies. http://cdimpact.org/publications/implementing-development-evaluations-under-severe-resource-constraints

Mahmood, S.S., Iqbal, M., Hanifi, S.M.A., Wahed, T. and Bhuiya, A. (2014). Are 'Village Doctors' in Bangladesh a Curse or a Blessing?' *BMC International Health and Human Rights* 10(18). www.biomedcentral.com/content/pdf/1472-698X-10-18.pdf (accessed 30 January 2015)

Mills, A., Ataguba, J.E., Akazili, J., Borghi, J., Garshong, B., Makawia, S., Mtei, G., Harris, B., Macha, J., Meheus, F. and McIntyre, D. (2012). Equity in Financing and Use of Health Care in Ghana, South Africa, and Tanzania: Implications for Paths to Universal Coverage. *The Lancet* 380.9837:126 33. www.sciencedirect.com/science/article/pii/S0140673612603572 (accessed 30 January 2015)

Mou, J.; Sian, M.; Griffiths, H.F. and Dawes, M.G. (2013). Health of China's Rural-Urban Migrants and their Families: A Review of Literature from 2000 to 2012. *British Medical Bulletin* 106:19-43. https://www.researchgate.net/publication/236924636_Health_of_China's_rural-urban_migrants_and_their_families_A_review_of_literature_from_2000_to_2012

Nikiema, B.; Haddad, S. and Potvin, L. (2012) 'Measuring Women's Perceived Ability to Overcome Barriers to Healthcare Seeking in Burkina Faso', *BMC Public* 12: 147, www.biomedcentral.com/1471-2458/12/147 (accessed 30 January 2015)

Omobuwa, O., Onayade, A.A., Olajide, F.O. and Afolabi, O.T. (2013). A Comparative Study of the Nutritional Status of Under-Five Children of Indigenous and Non-Indigenous Parentage in Ile-Ife, Nigeria. *Journal of Community Medicine and Primary Health Care* 25(1):1 11. www.ajol.info/index.php/jcmphc/article/view/95507/84849 (accessed 3 March 2015)

Paina, L. and Peters, D. (2012). Understanding Pathways for Scaling Up Health Services through the Lens of Complex Adaptive Systems. *Health Policy and Planning* 5: 365 73. http://heapol.oxfordjournals.org/content/early/2011/08/05/heapol.czr054.abstract

Radhakrishnan, R., Vidyasagar, S., Varma, D.M. and Sharma S. (2012). Design and Evaluation of Pictograms for Communicating Information about Adverse Drug Reactions to Antiretroviral Therapy in Indian Human Immunodeficiency Virus Positive Patients. *Journal of Pharmaceutical and Biomedical Sciences* 16(16):1-11. http://eprints.manipal.edu/76241/ (accessed 3 March 2015)

Richards, E., Theobald, S., George, A., Kim, J.C., Rudert, C., Jehan, K. and Tolhurst, R. (2013). Going Beyond the Surface: Gendered Intra-Household Bargaining as a Social Determinant of Child Health and Nutrition in Low and Middle Income Countries. *Social Science and Medicine* 95: 24   33. http://archive.lstmed.ac.uk/3367/

Schulz, Kenneth F and David A Grimes (2002). Blinding in randomised trials: hiding who got what. *The Lancet* 359: 696-700. http://apps.who.int/rhl/LANCET_696-700.pdf

Sundararaman, T. and Gupta, G. (2011). Indian Approaches to Retaining Skilled Health Workers in Rural Areas. *Bulletin of the World Health Organization* 89:73   77. http://www.who.int/bulletin/volumes/89/1/09-070862.pdf

Tilley, N. (2000). Realistic Evaluation: An Overview, paper presented at the *Founding Conference of the Danish Evaluation Society*. September. http://evidence-basedmanagement.com/wp-content/uploads/2011/11/nick_tilley.pdf (accessed 3 March 2015)

Varvasovszky, Z. and Brugha, R. (2000). How to Do (or Not to Do)... A Stakeholder Analysis. *Health Policy and Planning* 15.3:338   45. http://dess.fmp.ueh.edu.ht/pdf/Zsuzsa_Varvasovsky_2000_stakeholder_analysis.pdf

Wise, Jacqui (2001). Pfizer accused of testing new drug without ethical approval. *BMJ* 322(7280):194. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1119465/

Yahya, M. (2006) 'Polio Vaccines   Difficult to Swallow: The Story of a Controversy in Northern Nigeria', IDS Working Paper 261, Brighton: Institute of Development Studies, www.ids.ac.uk/files/Wp261.pdf (accessed 30 January 2015)

Zwarenstein, M. and Treweek, S. (2009). What Kind of Randomized Trials do we Need? *Journal of Clinical Epidemiology* 62: 461   63. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2679816/

Chapter 6: Ethical considerations in health systems research

Abbas Rattani
Meharry Medical College, School of Medicine
Adnan A. Hyder
Johns Hopkins Bloomberg School of Public Health &
Johns Hopkins Berman Institute of Bioethics

1.   Introduction

This chapter assumes a basic (introductory) familiarity with core terms in both health systems research and bioethics; for the latter these include the core principles of respect for persons (autonomy), beneficence, and justice (Coughlin 2008). We illustrate some of the issues that these principles are intended to address with two examples:

> Example 1: A measles immunisation programme for all children in a low-income country is to be created. The population density and local ecology lend itself to the proliferation of epidemics, and most children have had the infection by age three. It has been noted statistically that immunisations have the greatest impact on high mortality rates if administered by age one. Studies in high-income countries have shown that the immunological response to a measles vaccine is most effective at 15 months of age. Local research is needed to decide the ideal age at which the programme can produce the most impact on measles incidence and mortality.

> Example 2: A rapidly industrialising middle-income country has been expanding their transport and communication networks, with a large growth in healthcare facilities. The newly created hospitals and medical centres were potentially capable of providing complete coverage for the entire population. The Ministry of Health is concerned that despite the investment in health facilities and services, they are largely inaccessible to many individuals. A district health officer is appointed to improve the function of the health facilities. She discovers that a serious shortage of drugs and supplies at a medical centre was the result of the hospital siphoning off most of the drugs and supplies. She is to design a study to examine the misallocation of drugs within this system.

As noted in these two slightly modified cases from the World Health Organization (WHO) (Taylor 1984), it is apparent that research that addresses these problems - or *health systems research* - is different from other types of empirical clinical or public health research, and covers a wide range of subject areas that are focused on common health systems functions, such as stewardship, financing, resource inputs, and delivery of services (WHO 2009). Health systems research (HSR) is defined by the WHO as "the purposeful generation of knowledge that enables societies to organise themselves to improve health outcomes and health services" (WHO 2009:p7).

HSR is not usually research that focuses on the discovery or development of new interventions to improve health; rather, it is research that usually aims to understand *how* new interventions that are efficacious can be made more widely accessible to potential beneficiaries through policies, organisations and programmes (Gilson et al. 2011). While some HSR adopts the traditional randomised controlled trial (RCT) model (JPAL undated), many HSR studies are performed as non-randomised, controlled or non-controlled, prospective or cross-sectional assessments of new or modified health care programmes and strategies (Alliance for Health Policy and Systems Research 2012a). Given the macro focus of HSR, its participants and beneficiaries are often communities, hospitals, and healthcare institutions, as opposed to

individuals. Since HSR has its own definitions, methods, and analytic approaches, there is an increasing realisation that HSR raises ethical concerns that differ from those in other types of research; therefore, its ethical review should arguably be tailored to address the features and unique ethical challenges that are particularly salient (though not exclusive) to HSR.

Unfortunately, many (if not most) institutions often use the same review criteria and review processes for HSR studies as for clinical trials, which can potentially create an imprecise application of criteria, confusion on the part of research teams, and unnecessary delays. Currently, it is not clear whether institutions are equipped to adequately address and ethically evaluate HSR in their research ethics committees (REC) or institutional review boards (IRB) (Bachani et al. 2016, Hyder et al. 2015, Hyder et al. 2012a). This is especially true for HSR in low- and middle-income countries (LMICs), where this research often plays a critical role in efforts to strengthen health systems and improve healthcare delivery.

Based on the presumption that certain kinds of ethical issues may be particularly relevant to and salient in HSR, this chapter explores several of these issues. We outline eight areas of ethical relevance that are particularly salient in HSR (though not unique to HSR) that may require special attention during ethics review, especially in LMICs. This set of issues is used to demonstrate only some of the salient features that might arise during HSR - they are not exhaustive and readers are encouraged to add to the list below.

2.   Type of research subjects in HSR

The 'research subjects' in HSR studies can either be humans or non-humans, each with their own respective ethical challenges.

> Example: A team of engineers in collaboration with public health workers is designing a vehicle crash reduction study on a new technology created to improve driver and passenger safety in auto rickshaws.

Non-human 'subjects' in HSR may include units of interventions, such as motorcycles equipped with safety features in a vehicle crash reduction study, or units of allocation, such as hospitals or schools involved in a study of cost containment budgetary strategies. When reviewing study protocols with these kinds of non-human subjects, it may be challenging for RECs to assess what role various actors should play in the authorisation and implementation of the study. For instance, when the intervention involves safety features for products, how should RECs weigh the interests of the manufacturers as well as consumers of these products? When schools or hospitals are the unit of allocation, how should the teachers and students in these institutions factor into the ethical analysis? For HSR studies, the level of impact goes far beyond individuals, and even research with non-human 'subjects' requires consideration of a wide range of stakeholders who may be affected by the investigation. As of yet, there is no standard universal guidance on how to assess and balance the interests of these various parties in HSR studies.

When considering human subjects, HSR studies may target individuals as units of allocation or intervention, though more commonly they are directed at groups of people, as in population-level or cluster-based studies. The emphasis on groups of people as research subjects introduces the ethical challenge of defining the moral status of a group or community as opposed to individual persons. Identifying appropriate representatives or leaders of these groups may also be trying, especially when assessing their legitimacy and source of authority. The principle for respect for communities has been proposed as a means of defining moral worth and protecting the interests of a given community (Weijer and Emanuel 2000). This principle understands the community as a source of values, a social structure that sustains its members and makes decisions for its members. Therefore, RECs concerned with respect will

have to think far beyond the typical construction of respect for persons, concerned with individuals (and often focused on consent), and instead adopt the broader interpretation of respect for communities to determine what is required. Their review of such an HSR study will have to be considerate of study community priorities and norms, and determine appropriate levels of engagement with local leadership, which presents further challenges, particularly in pluralistic communities embodying a range of diverse interests. This extension of 'respect' from an individual to a population requires further exploration for global health research (Wallwork 2008).

HSR studies involving large populations or groups of people also require a broader interpretation of burdens and benefits and how they may be differentially distributed across various study populations. Similarly, concerns around potential harms need to be reviewed, such as a group reputation potentially affecting individuals - for example, a hospital that is perceived to provide low-quality care may develop a reputation that affects the flow and type of individual patients who visit it. This concept has been described as *group harm* by which 'members suffer it *by virtue* of their identification with or participation in the group' (Wallwork 2008). This presents complications for RECs in assessing benefits and harms at the community level. The current norm for reviewing research focuses on the individual, but this narrow application of principles at the individual level is not well suited for assessing HSR, in which group-level interventions and impacts require a much broader lens.

3.   Informed consent in HSR

Consent can be obtained similarly in both HSR and clinical studies in the event of individuals receiving a particular intervention. However, in many HSR studies where interventions are administered to an entire group, the consent process has to involve authorisation at multiple levels, engaging community or institutional leaders as well as affected individuals.

> Example: With the incidence of malaria growing in a city in Burundi, the Ministry of Health approved an intervention designed by researchers in Sweden to alter standard procedures on malaria prevention and control. Seeking informed consent from individuals would be impractical, and so the research ethics committee and the Ministry of Health opted for group consent from each village.

In some studies where the intervention can be delivered at an individual level, such as with malaria bed nets, researchers may require consent from both the community leadership, as well as from individuals or households participating in the study. However, other interventions, such as adjustments to standard procedures or drugs offered at public facilities, broadly impact a large number of people for whom obtaining consent would be impracticable. Alternatively, individual informed consent in HSR studies that focus on area-wide interventions such as spraying for malaria control or building speed bumps for road safety, may not be possible.

In these instances, group consent (or permission) is usually obtained through representatives and often paired with community outreach and education. In some circumstances, participants still have the ability to opt out and can take voluntary actions to exclude themselves from study participation (for example avoiding public facilities or seeking private providers). Since in many types of HSR (and this includes cluster-randomised trials of certain group interventions), individual informed consent may not be obtainable, some have argued that ethics committees have an obligation to ensure that the justification for waiving consent is adequate (Sim and Dawson 2012, Taaljard et al. 2009, Weijer et al. 2011). A trade-off may need to occur in decisions regarding the choice between individualised consent and ability to conduct valid HSR studies; indeed, if the societal value of the HSR study is high enough, it may allow concerns of greater benefit to outweigh individual autonomy concerns and permit practical studies to move forward (Hutton et al. 2008).

Consent involving groups of people may not be specific only to HSR, but is starting to become a 'norm' in many HSR studies in LMICs. Questions persist about how groups should be defined and how formal permissions and consent processes are being administered in LMICs. Key to this is addressing the issues of group representation, legitimacy of representatives, authority structures, and coverage of the consent process. For example, concerns have been well documented in the literature around the validity of leaders who give consent for a group; potential exclusion of vulnerable groups including women; and ability of individuals within the groups to opt out (Cassell and Young 2002, Davis 2000, Diallo et al. 2005, Emanuel et al. 2004, Ijsselmuiden and Faden 1992, Weijer and Emanuel 2000). Thus, consent from groups and/or representatives is often necessary and yet case-by-case discussions are needed to determine whether it is sufficient.

An important issue is that of defining subjects for consent irrespective of whether they are individuals or groups. For example, common requirements for informed consent may not apply to many HSR studies. In the United States (US), informed consent applies to 'research subjects' defined as those actively involved in research (Protection of Human Subjects Research 2009). However, in circumstances where no 'direct' subjects are identifiable, as is the case in many HSR studies, is such a requirement for consent appropriate? For some HSR projects, the lack of identifiable human subjects and aggregation of data for analysis may lead IRBs and RECs to designate these studies as 'non-human subjects research', which would exempt them from the consent requirements specified in the US regulations.

Furthermore, the US federal regulations also waive consent when studies fulfil four conditions: the research is no more than 'minimal' risk; the rights/welfare of subjects are not adversely affected; the research cannot be carried out in other ways; and the subjects will be debriefed (when appropriate) (Protection of Human Subjects Research 2009). Applying these conditions to HSR studies would mean that many of them could obtain a waiver of consent. The nature of group interventions, which often lack identifiable direct subjects and are built into health systems responses, makes HSR studies amenable to such waivers. Appropriate ways to handle consent, authorisation, and authentic community engagement for group-level interventions characteristic to HSR remain a challenging area for investigators and ethical review boards.

4.   Units of intervention and observation in HSR

Unlike typical clinical research, in which interventions are often administered to individuals who are then observed for potential effects, HSR often targets a unit of intervention at a more macro level and then assesses its impact at a more micro unit of observation. In other words, the units of intervention and observation are often not the same.

Example 1: Health systems researchers have designed a study to provide local taxi drivers with incentive payments to transport pregnant women to the clinic for antenatal care and delivery. Although the intervention is administered to the taxi drivers, the outcomes data are being collected on mothers and infants within the intervention community.

Example 2: A hospital has recently decided to introduce quality assessment activities for infection control by teams of health providers. However, the hospital plans on collecting outcome data on hospital-acquired infections among patients admitted to those hospitals.

In the first example, the local taxi drivers were the unit of intervention, while the outcomes data were collected on mothers and infants, which would be the unit of data collection/observation.

The use of different units for intervention and observation creates a new set of challenges for ethical review. One issue is in terms of defining and assessing risks and benefits for multiple levels of research participants: the research subjects who might be the unit of intervention (sometimes called primary), and other research subjects from whom data is collected (sometimes called secondary). In the second example, the teams of health providers are the unit of intervention, and the outcome data are collected from hospital-acquired infections among patients, which is the unit of observation. Furthermore, the hospital staff (doctors, nurses) and patients would all be research participants. How should RECs assess the study with appropriate regard for *all* groups of research subjects whose well-being can be impacted by the intervention?

This also raises important questions for the consequential targets and nature of informed consent; that is, who should be involved in the informed consent process and decide when individual consent of some (secondary) participants might be impracticable, and what should be the standards for informing them of the study? If data collection involves a measurable burden for some participants, such as additional interviews, does this incremental burden necessitate greater participation in the consent process? It is clear that having different units of intervention and data collection presents unique challenges for how practical matters of risk benefit analysis and informed consent are carried out for HSR studies.

5.   Risk assessment in HSR

Risk assessment in HSR is considered an area with serious practical and ethical challenges for HSR in many contexts (Peters *et al.* 2009). Traditional risk assessment for clinical research studies focuses on physical risks to participants, with some additional attention to psychological and social risks associated with participation. However, the types of risks associated with HSR studies can be quite different from clinical research, often with the largest risks manifesting in social, financial, or communal harms.

> Example 1: To reduce the incidence and prevalence of smoking in Mumbai, India, a health systems research group wants to use social media as their intervention for smoking prevention. A member of the research group is concerned that a social media campaign against smoking could overtly stigmatise current smokers or the message could get inaccurately modified somewhere in the communication chain, proliferating harmful misinformation.

> Example 2: A team of researchers at a prominent university in Uganda plan on designing a programme to provide conditional cash transfers as incentives to pregnant women to deliver their children in a hospital, arguing that institutional newborn delivery results in better outcomes. A health economist at the university advises that the incentives for women to deliver in a facility could expose participants to a variety of harms in places where home birthing is the norm, not to mention the potential of the cash transfers to be a more macro threat by distorting local economic markets.

While the use of sound and appropriate designs to minimise risk still applies in HSR studies, different approaches might be needed for both assessment and mitigation. As noted in the above examples, identifying and quantifying risks in an HSR study on using social media for smoking prevention in a population or the use of financial incentives for promoting institutional newborn deliveries (conditional cash transfer) requires a much more in-depth understanding of the underlying social conditions and system-level factors.

The issue of risks also relates back to appropriate modes of consent. In typical clinical research, participants are directly informed of the potential risks, and by consenting they express their willingness to accept these risks as part of their participation. There are many

HSR designs in which individuals may not have this opportunity to directly consent to the exposure to risks associated with the study. Further concerns arise when potential risk levels vary across subsets of the population group, especially when the local leadership granting authorisation for the research may not represent these subgroups. One could imagine communities in which a practice under investigation might go against the norms of a religious or cultural minority or some study objectives may disproportionately burden the extremely poor. When risks are evaluated at an aggregate level across the population and marginalised groups are not represented in decision-making, the potential for undue burden and disregard for these subgroup values have clear ethical implications related to distributive justice and respect for persons.

Although many HSR studies are typically classified as low risk, a risk benefit analysis remains important and requires broader interpretation of how harms may result. Some of these present new challenges in defining 'minimal risk,' since knowledge of negative group characteristics might pose social concerns in how a health system treats members of that group. Moreover, defining who is at risk, inclusive of all types of research subjects, varies in HSR and may include several stakeholders involved in a study, such as providers, recipients, beneficiaries, observers, institutions, and tribes. Considerations for risk assessment therefore have to go well beyond a simple focus on individual participant concerns in HSR studies. Additionally, monitoring systems would need to be set up to report adverse consequences resulting from the research so that these harms are appropriately captured during implementation.

6.   Defining benefits, beneficiaries, and fair benefits in HSR

Research subjects in LMICs do not always have access to the same standard of care enjoyed by subjects in high-income countries. Establishing a standard of care becomes difficult with varying types of health systems that are often the context (and the object) of HSR studies. Hence, notions of 'best care available', which have been promulgated in research ethics guidelines, may not be relevant if they are applied to LMIC health systems.

> Example: A researcher in Bangladesh has designed a study that examines health systems issues within his city. His study protocol calls for referring patients/participants to their local facility for receipt of appropriate care, but due to health systems inefficiencies, the quality of these facilities or the standard of care available may not be equivalent. He knows that in his application to the research ethics committee, though this may seem like equivalent treatment of patients using different facilities, variation between those facilities will mean that in reality there may be substantial disparities.

Arguably, the very concept of standard of care continues to remain ambiguous (Hyder and Dawson 2005, London 2000). This ambiguity results in challenges in assessing the implications of opposing standard of care arguments, in recognising important differences in their supporting rationales, and even in identifying the major source of disagreement (London 2000). For example, others have attempted to address the standard of care debate from a health systems perspective, arguing that the structure and efficiency of national health systems have been neglected in arguments about the standard of care in research (Hyder and Dawson 2005).

Addressing the current global variability is a challenge, especially in elucidating benefits, beneficiaries, and the range of responsibilities in offering benefits to participants in global health systems research, particularly in satisfying the requirement of research to provide social value. One ongoing debate is whether the individual participants in a study or the communities from which they are drawn should be counted as the beneficiary, with implications for what is due to each during the course of the study (for example benefits like capacity building) and after the trial concludes (for example post-trial access and benefit sharing) (Lairumbi et al.

2011). This conversation reflects the current bias in research ethics literature to consider the individuals enrolled in studies as the primary participants and beneficiaries. However, because the goals of HSR are to make improvements at the systems level, and units of intervention in HSR are groups, with individuals as indirect beneficiaries, this dialogue about what is due to individuals versus broader communities is more important for HSR. Few guidelines discussing beneficiaries of research include the 'larger community/host country', further highlighting how one of the main beneficiaries in HSR may be under-recognised when applying these guidelines for review of HSR studies.

Several international and national ethics guidelines support provision of diverse types of benefits during and after studies, yet many of the benefits in HSR may be left out, such as improvements in healthcare delivery systems, actual provision of treatment, human and material capacity building, and health systems strengthening. It is also important to regard more equitable distribution of existing resources as benefits in HSR; this means that addressing inequities in health provision is another form of benefit often considered in HSR studies, especially those that work on larger communities or countries. As a result, it appears that commonly used international and some national research ethics guidelines might not be addressing the forms and types of benefits in HSR or the beneficiaries of HSR and, thus, their usage by ethics committees poses challenges for review of HSR studies.

7.   Nature of interventions in HSR

Ethical challenges that are intervention-specific in HSR vary from concerns around scientific rationale to distribution of benefits to sustainability issues. For instance, in the case of HSR testing new delivery methods (for example for child health), one could question whether there is appropriate evidence to support the testing of a new approach or challenge the need for innovation over continuing with existing delivery systems (such as community health workers versus facility-based delivery).

> Example: The most common cause of newborn mortality is preterm birth. A local community in a low-income country has been using the 'kangaroo mother care' intervention for preterm infants weighing less than 2kg, which includes skin-to-skin contact, support of the relationship between mother and child, as well as exclusive and frequent breastfeeding. This form of care has been shown to reduce infant mortality in some hospital-based settings in low- and middle-income countries. A community health worker and researcher wishes to use the population-level experience as a proof-of-concept to undergo a large-scale trial of testing this new method of delivering care in Zambia.

Where there is not much prior evidence on an approach, are theory and hypothesis enough to justify testing the intervention, or is there some population-level experience that should be required to demonstrate proof-of-concept prior to a larger-scale trial of the new method as noted in the example above?

This is of particular concern in LMICs, since their need for novel interventions to deliver services efficiently makes them arguably ideal candidates for testing health systems innovations, and if there is meagre evidence supporting the effectiveness of interventions, these resource-constrained settings may bear a disproportionate burden in the generation of global health systems innovations. Implicit in this concern are (1) the obligation of researchers to not impose undue harm upon populations, which may occur in the absence of sufficient evidence (for example distortion of a local health market), and (2) issues of distributive justice, in which disadvantaged communities assume the risks of research on interventions that will ultimately benefit more advantaged populations - an increasing concern as more high-income nations adopt innovative models from developing country settings (Fry et al. 2011).

Similarly, another ethical concern is the potential for harm with new health delivery methods and associated safety issues. The kinds of harms resulting from HSR tend to be more obscure, downstream, and harder to quantify than those typically associated with a clinical study. In order for RECs to adequately assess the potential harms associated with certain types of HSR, they will have to rely on the existing evidence base of the approach, with a good understanding of the history of a particular delivery method and its success or failure with similar types of health interventions. However, for many novel approaches, there might be insufficient prior evidence available to inform the ethical review process.

A critical concern with HSR is that it can also blur the distinction between research and non-research processes. For example, it is important to make the distinction between quality improvement (QI) projects, which are meant to improve service deliverance and process performance, and research, which is meant to produce generalisable or transferable knowledge. Though the former is typically exempt from ethical review, pertinent issues may overlap for both QI and HSR, regardless of what may be legally required *vis-à-vis* regulations. From a practical standpoint, this range and variability can pose difficulties for RECs in gaining experience reviewing certain kinds of HSR and applying recommendations consistently. As compared to clinical trials, which often share common features and have more clearly identified areas for ethical consideration established in the literature, HSR may present unique challenges for review committees with each new protocol. This is especially the case when HSR refers to areas wherein the REC does not have much experience, which can impact the quality of the review and further strain the limited capacity of RECs to assess study proposals in a thorough and efficient manner.

Finally, in LMICs where a lack of access to health interventions exists, ethical concerns around future availability become salient. Will the community involved in the trial continue to have access to beneficial services provided as part of the study? While the issue of post-trial access is not unique to HSR and has been widely discussed in research ethics literature, this issue is of particular import for HSR given the well-documented lag in, or absence of, research-to-policy translation ([Grady 2005](#), [Lavery 2004](#)). What impact might the temporary change in health delivery mechanisms or available services have (during a study) on the community, and could this disruption in the status quo have net negative consequences for the population? At the systems level, decisions to adopt new approaches for providing health to the population often weigh costs and benefits at the aggregate level, so even interventions that show improvements for those involved in the study may not be taken up in the end if they do not prove cost-effective. These concerns must play a role in how local and national health sectors analyse and respond to the results of an HSR study and raise questions about what obligations exist for research institutions and funders conducting such work.

In sum, HSR is fundamentally about translating efficacious interventions into effective practice at the population level. As a result, the interventions under investigation in HSR can vary greatly, as can their methods of delivery, resulting in ethical issues quite specific to a given study. These interventions might be health messages, incentives, measurement tools, performance guides, intervention packages, financial subsidies, or delivery systems. Therefore, typical interventions in HSR can involve new methods of delivery or dissemination of existing or proven interventions, novel approaches for creating demand for efficacious interventions, new packaging of two or more interventions for enhanced programme effectiveness, or knowledge generation on costs or cost-effectiveness for policy impact. This diversity in the intensity, invasiveness, and duration of implementation requires a very good understanding of the intervention in each HSR study in order to define relevant ethical issues.

8.   Appropriate controls and comparisons in HSR

The nature of control groups can vary in HSR studies, and the ways groups are compared are often not consistent with common clinical research study designs, such as placebo-controlled studies, where the 'gold standard' involves comparing an 'intervention' group with a 'non-intervention' group. For instance, if an HSR study is testing a new delivery method for a proven intervention, then the comparison group may have an older delivery method, or if an HSR study is testing a new package of existing interventions (say A and B together), then the comparison group may receive them separately (either A or B alone). The selection of these comparison locations is also often not done randomly, but rather by systematic matching or even geographical or logistical convenience. As a result, comparison groups in HSR studies pose challenges to the ethical review process when these control groups receive different types of interventions; and there is a wide variation of possibilities in what might constitute comparison groups.

> Example: A study to evaluate the efficacy of a new health safety curriculum in local medical centres is underway in Dodoma, Tanzania. Participants from the intervention group share their knowledge with members of the control group via social networks and staff transfers. The control group's integrity is effectively compromised and the extent of the 'contamination' is difficult to assess and threatens to undermine the interpretation of the magnitude of the findings.

HSR presents challenges for establishing appropriate comparison groups. As compared to clinical trials, it is more difficult in HSR studies to control for a variety of extraneous variables that could impact results. This is due to the fact that HSR often involves interventions that take place within existing, real-world settings, while clinical trials occur in highly controlled experimental settings. Therefore, many (especially low-cost) HSR studies use comparators of convenience, such as data from similar districts or cities, or quasi-experimental pre-post designs, often applying complex statistical techniques in an attempt to account for non-parities or temporal confounders. In order to ensure the internal validity of these studies - a necessary ethical requirement of all research - RECs should be equipped to evaluate the techniques used in HSR to determine if studies have adequately controlled for the challenges of imperfect comparison groups (Emanuel et al. 2004). This will have implications for the future applicability of the study findings and their social value, in addition to ensuring respect for the communities participating in the HSR study.

In addition to determining who should serve as the control, there is also the question of what should be provided to the control groups. Although ethical debates concerning the appropriate use of placebos versus active controls are not exclusive to HSR and have been ongoing in the literature for many years surrounding both clinical and implementation trials, these concerns are particularly acute in the context of HSR in LMICs (Emanuel et al. 2000, Emanuel and Miller 2001, Freedman 1990, Miller and Brody 2002). If there is little evidence available concerning the effectiveness of current systems of practice, it becomes difficult to choose what to test a new health system approach or combination of approaches against (if anything), whereas in clinical investigations testing equivalency or superiority, there is often a much more robust evidence base about the current standard of practice. Furthermore, where an HSR study seeks to assess packages of multiple beneficial interventions that have potentially synergistic effects, what subset(s) of these interventions should be provided to the control group(s)? If the researchers are seeking to find the most cost-effective package of services to produce the desired health impact, they must balance their obligation to provide existing beneficial interventions to their participants against their aim to produce information for evidence-based policy that will ultimately provide the greatest societal benefit.

Another relevant factor for many cluster-based studies arises when they use a staged introduction (or stepped wedge design), in which the intervention is rolled out sequentially to participating groups or clusters so that even the control groups receive the intervention by the end of the study. While staged roll-out is often considered to be more ethically acceptable than providing no intervention to control groups, there is still the risk that the control communities will feel unfairly disadvantaged. This could pose validity threats due to varying external conditions over time or contamination from neighbouring clusters via information diffusion, and may also raise issues of justice and fairness for the clusters receiving the intervention so much later than their counterparts (Brown and Lilford 2006). These types of specific issues must be understood within the overall aim of HSR - to inform real-world practice and produce social value. In the interest of good science, RECs must be better equipped to evaluate these options and determine whether HSR studies have adequately considered appropriate comparison groups (Emanuel et al. 2004).

9.   Inclusion of vulnerable groups in HSR

As the volume of research in LMICs increases, the role (and protection) of the highly vulnerable (for example women or stigmatised groups) in research among the poor or generally vulnerable groups becomes a serious challenge.

> Example: A Malawi HIV clinic has created a programme to incentivise HIV testing and collection of test results. Recently, self-identifying gay and lesbian individuals, a highly stigmatised and vulnerable group in Malawi, were seen entering the clinic and were later beaten by an unidentified mob.

These especially disadvantaged groups are often left out of general improvements in healthcare due to lack of access or lack of power, and become further marginalised. For instance, in locales where freedom of movement is restricted for women, their access to basic health services may be limited. Therefore, improvements in the delivery of services at health centres may not translate to benefits for this subgroup. Furthermore, as seen in the above example, interventions aimed at stimulating demand for services may overlook the social or cultural risks to individuals if they pursue these services. Thus, including these concerns for particular vulnerable subgroups who face acute risks and whose position may not be represented in many models for group authorisation is an important consideration that needs special attention when evaluating risks and benefits associated with HSR. It is uncertain how well RECs in general are equipped to address the specific concerns that these highly vulnerable subgroups pose in a study. This is an increasing challenge in addressing the ethical issues of conducting much-needed HSR in LMICs and remains largely unexplored.

HSR often involves vulnerable populations, especially in LMICs, where the general population's impoverished condition may already place them at historical disadvantage. This type of vulnerability raises ethical concerns around risks of exploitation, coercion, and abuse. The *International Ethical Guidelines for Biomedical Research Involving Human Subjects* by the Council for International Organization of Medical Sciences (CIOMS) specifically state: "Special justification is required for inviting vulnerable individuals to serve as research subjects and, if they are selected, the means of protecting their rights and welfare must be strictly applied" (Guideline 13) (Council for International Organizations of Medical Sciences 2002).

However, many HSR studies, especially in LMICs, are in fact conducted with the primary aim of reaching vulnerable groups and providing access to existing or proven interventions for those communities. When such groups are the focus, the assessment of risks associated with these vulnerable populations continues to remain a challenge, since it is also ethically important to try out new ways of delivering and accessing care in the same population to have relevance. Paternalistic protection of vulnerable groups from HSR might compromise the

opportunity to find solutions to some of the most important health system challenges. One characteristic challenge in the ethical review process of HSR is identifying when it is acceptable to pilot health systems innovations intended for broad scale-up among particularly vulnerable groups, who may realise the most benefit but who may, conversely, be subject to further harm as systems researchers explore new techniques.

Vulnerable populations suffer and face the worst burden of health due to system weaknesses, reinforcing the need to emphasise the larger notion of fairness (Daniels 2006). Fairness is an important consideration in the ethical review of HSR, especially as it relates to communities and populations that may become vulnerable, not because of inherent weakness, but because of the context in which they are operating (Bamford 2014, Hurst 2014, Hyder et al. 2014). Thus, on the one hand, HSR responds to such lack of fairness by trying to identify strategies to reduce inequalities; however, at the same time, the conduct of HSR can affect fairness. Therefore, RECs need to evaluate this potential impact for each proposed study.

10. Conclusion

There are several limitations to the conceptual exploration above that are worth considering. First, the definition of HSR varies depending on the type of research, location, or source considered and makes consideration of this field challenging. However, a unified definition is necessary, and global meetings are now focusing on further defining and enhancing the field (Alliance for Health Policy and Systems Research 2012a; 2012b; Global Forum for Health Research 2004). Second, there are activities that are often in the grey zone between research and non-research and that can be considered part of the HSR agenda in LMICs. Two examples include: (1) quality assurance methods, which are used for performance management and research; data may be collected to inform practice only or become part of formal research activities in HSR (Heiby 1993, 1998, Reinke 1995, Zeitz et al. 1993); and (2) public health surveillance activities, which are not traditionally considered research but may be part of HSR for example, where a new disease surveillance system is being pilot tested for the first time (Lee and Thacker 2011; Lee et al. 2012). These types of approaches, when used in HSR, can add further complexity to ethics considerations. As a result, the mutually exclusive categorisation of HSR as either research or non-research by ethics committees and current guidelines is a source of challenge for the field.

Third, a discussion focused on teasing out differences between HSR and other research tends to downplay the many similarities across all types of health research; in many instances the differences are less stark and similarities more common. However, a conceptual exploration has to use some real-world generalisations that can stand merit even though specific exceptions can be defined.

Fourth, the diversity of HSR extends beyond the typical examples that have been provided, and can include the conduct of long-term HSR in the same site, such as the use of HSR in demographic surveillance sites across LMICs. Such longitudinal and often long-term HSR (years or decades in the same sites) can lead to different types of ethics issues associated with more dynamic concerns (Hyder et al. 2012b). Important conversations and areas for further exploration for HSR ethics include vulnerable populations, big data, ancillary care obligations, distribution of responsibility, and the potential (and possible moral obligation) of health systems research to help reduce health disparities between and within countries (Bamford 2014, Dereli et al. 2014, Gupta 2014, Hurst 2014, Hyder et al. 2014, Olson 2014, Pratt 2014, Rennie 2014).

Since health systems research, especially in LMICs, is substantively different from other types of research with its own set of objectives, approaches, methods, and analytic goals it warrants special or nuanced considerations in its ethical review. Some of these ethical concerns may

be more salient than the usual ethics review of other types of research such as clinical research (Annex Table 1). An ethics review of HSR that uses exactly the same criteria and ethical analysis as for clinical research may place an overemphasis on features that are not particularly relevant in HSR, and may not adequately capture the unique kinds of benefits and risks present in HSR. Thus, untailored review can result not only in practical inefficiencies, but also in unjustified research activities and inadequate protection of participating communities and individuals.

Ethical review of HSR does not always fit with the existing review paradigm born from the typical clinical research setting (Hyder et al. 2014, London et al. 2012). Additionally, more exploration is needed to understand the possible breadth of ethics issues that may apply to HSR in various contexts, and there is much to be learned from overlapping disciplines that have particular relevance to larger HSR concerns, such as health systems transformations (Daniels 2006). HSR studies ought to reflect fair terms of social cooperation between communities and researchers, be relevant to the health needs of the host communities, and have a favourable risk benefit ratio (Emanuel et al. 2004). Such responsiveness to host communities helps form collaborative partnerships in which all stakeholders (participants, researchers, brokers) are considered moral equals of each other. These concerns are important for HSR, as research resources themselves can have a direct impact on the distribution of opportunities in a community related to jobs, training, placement of facilities or site selection, with implications for distributive justice and fair equality of opportunity. This discussion can even be extended to include certain public health ethics obligations discussed in the literature, such as social duty, reciprocity, solidarity, stewardship, trust, and accountability (Baum et al. 2007, Swain et al. 2008, Thompson et al. 2006, Upshur 2002).

HSR is necessary to ensure health systems strengthening, quality of care, and evidence-informed public policy creation. HSR researchers must carefully define their intent and goals and openly clarify the values that may influence the premises and design of their protocols. In order to have appropriate ethical review of HSR, there is a need to have a deeper understanding of how to apply traditional ethics review criteria in ways that are relevant to the features of HSR, and further guidance to researchers and reviewers addressing the broader issues arising in the context of systems-level interventions.

Annex

Table 1: Ethics considerations of special relevance in HSR in LMICs

Ethics issue	Application to health systems research in LMICs
Nature of interventions	System-based such as delivery systems, financing, human resources, or policies
Type of subjects	Groups of people or communities
Units of intervention and observation	Often different such that intervention is distributed to one group and measurement is based on another
Informed consent	Group consent and permissions needed (in addition to individual consent)
Comparison groups	Comparators often receive different interventions or are observed in the real world
Risk assessment	Broad range and different types of minimal risk - social, communal
Inclusion of vulnerable groups	The focus of HSR and proposed beneficiaries
Benefits assessment	Expanded definition including training, infrastructure, health systems strengthening

Some questions to promote thinking about ethics issues

1. Take one of the eight ethical issues identified above and list three reasons why you: (1) agree that it is different for health systems research; and (2) think it is similar to other types of research.

   Probe: Then list three ways in which you feel this specific ethics issue can be addressed by health systems researchers.

2. You are about to start a health systems research study in a district of Uganda with 30 villages. The study will train community health workers in 15 villages on child health in year 1 and then provide the same training to workers in the other 15 districts in year 2. The study will monitor childhood diseases in all villages for two years. Describe three ethical concerns you might have in this study. Who do you think should give consent for the study in the district?

   Probe: What risks is the population being exposed to and how would you manage them if you were the study director?

3. What other ethical concerns (apart from the eight above) can you think of that may be particular to health systems research that differentiate it from clinical research?

   Probe: And do you know of other ethical frameworks within public health that might address these other ethical concerns?

4. What counts as a benefit in health systems research and which benefits are due to communities versus their members?

   Probe: How can this guide research designs and research ethics committees' decisions about the obligations of health systems researchers to participating groups and individuals during and after a trial?

5. Can you identify actual health systems research studies wherein ethical considerations may have been overlooked by the current review process?

   Probe: Do a PubMed search and review some studies.

References

Alliance for Health Policy and Systems Research (2012a). *Health Policy And System Research: A Methodology Reader*, (ed. L. Gilson). Geneva, Switzerland: World Health Organization. www.who.int/alliance-hpsr/alliancehpsr.reader.pdf (accessed 12 March 2015)

Alliance for Health Policy and Systems Research (2012b) 'Global Symposium on Health Systems Research', World Health Organization, www.who.int/alliance-hpsr/hsr-symposium/en/ (accessed 12 March 2015)

Bachani, Abdulgafoor M., Abbas Rattani, Adnan A. Hyder (2016). A scoping study on the ethics of health systems research. Developing World Bioethics. doi:10.1111/dewb.12117. http://onlinelibrary.wiley.com/doi/10.1111/dewb.12117/abstract

Bamford, R. (2014). Ethical Review of Health Systems Research: Vulnerability and the Need for Philosophy in Research Ethics. *American Journal of Bioethics* 14(2):38-40. http://www.tandfonline.com/doi/abs/10.1080/15265161.2013.868954?journalCode=uajb20

Baum, N.M., Gollust, S.E., Goold, S.D. and Jacobson, P.D. (2007). Looking Ahead: Addressing Ethical Challenges in Public Health Practice. *Journal of Law, Medicine, and Ethics* 35(4):657-667. https://www.researchgate.net/publication/5771112_Looking_Ahead_Addressing_Ethical_Challenges_in_Public_Health_Practice

Brown, C.A. and R.J. Lilford (2006). The Stepped Wedge Trial Design: A Systematic Review. *BioMed Central Medical Research Methodology* 6:54. http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-6-54

Cassell, J., and A. Young. (2002). Why We Should Not Seek Individual Informed Consent for Participation in Health Services Research. *Journal of Medical Ethics* 28: 313-17. http://jme.bmj.com/content/28/5/313.full.pdf+html

Coughlin, Steven S. (2008). How Many Principles for Public Health Ethics? Open Public Health Journal 1: 8-16. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2804997/

Council for International Organizations of Medical Sciences (2002) *International Ethical Guidelines for Biomedical Research Involving Human Subjects,* Geneva, Switzerland: Council for International Organizations of Medical Sciences, www.cioms.ch/publications/layout guide2002.pdf (accessed 12 March 2015)

Daniels, N. (2006).Toward Ethical Review of Health System Transformations. *American Journal of Public Health* 96(3):447-51. https://www.hsph.harvard.edu/benchmark/pdf/Daniels_Toward_Ethical.pdf

Davis, D.S. (2000). Groups, Communities, and Contested Identities in Genetic Research. *Hastings Center Report* 30(6): 38-45. http://onlinelibrary.wiley.com/doi/10.2307/3528452/full

Dereli, T., Coşkun, Y., Kolker, E., Güner, O., Ağırbaşlı, M. and Ozdemir, V. (2014). Big Data and Ethics Review for Health Systems Research in LMICs: Understanding Risk, Uncertainty and Ignorance - and Catching the Black Swans? *American Journal of*

Bioethics 14(2):48-50.
https://www.researchgate.net/publication/260168025_Big_Data_and_Ethics_Review
_for_Health_Systems_Research_in_LMICs_Understanding_Risk_Uncertainty_and_I
gnorance-And_Catching_the_Black_Swans

Diallo, D.A., Doumbo, O.K., Plowe, C.V., Wellems, T.E., Emanuel, E.J. and Hurst, S.A.
(2005). Community Permission for Medical Research in Developing Countries.
Clinical Infectious Diseases 41(2):255-59.
http://cid.oxfordjournals.org/content/41/2/255.full.pdf

Emanuel, E.J., Wendler, D., Killen, J. and Grady C. (2004). What Makes Clinical Research in
Developing Countries Ethical? The Benchmarks of Ethical Research. *Journal of
Infectious Diseases* 189(5):930-37.
http://jid.oxfordjournals.org/content/189/5/930.full.pdf+html

Emanuel, E.J. and Miller F.G. (2001). The Ethics of Placebo Controlled Trials - A Middle
Ground. *New England Journal of Medicine* 345(12):915-19.
https://www.unige.ch/medecine/ieh2/files/1014/3472/9160/me-7-Emanuel-Placebo-
research.pdf

Emanuel, E.J., Wendler, D. and Grady, C. (2000). What Makes Clinical Research Ethical?
Journal of the American Medical Association 283(20):2701-11,
http://www.dartmouth.edu/~cphs/docs/jama-article.pdf

Freedman, B. (1990). Placebo-controlled Trials and the Logic of Clinical Purpose. *IRB:
Ethics and Human Research* 12(6):1-6.
https://www.jstor.org/stable/3564541?seq=1#page_scan_tab_contents

Fry, C.V.; Marjanovic, S.; Yaqub, O. and Chataway, J. (2011). *Health Innovation Transfer
from South to North*. Santa Monica: RAND Corporation.
www.rand.org/pubs/documented briefings/DB616.html   (accessed 12 March 2015)

Gilson, L., Hanson, K., Sheikh, K., Agyepong, I.A., Ssengooba, F. and Bennett, S. (2011).
Building the Field of Health Policy and Systems Research: Social Science Matters.
PLoS Medicine 8(8): e1001079.
http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001079

Global Forum for Health Research (2004). *Strengthening Health Systems: The Role and
Promise of Policy and Systems Research*. Geneva, Switzerland: Alliance for Health
Policy and Systems Research. http://www.who.int/alliance-
hpsr/resources/Strengthening_complet.pdf

Grady, C. (2005). The Challenge of Assuring Continued Post-Trial Access to Beneficial
Treatment. *Yale Journal of Health Policy, Law, and Ethics* 5(1): 425-35.
http://digitalcommons.law.yale.edu/yjhple/vol5/iss1/15/

Gupta, S. (2014). Ethical Review of Health Systems Research in Low- and Middle-income
Countries: Research Treatment Distinction and Intercultural Issues. *American
Journal of Bioethics* 14(2):44-46.
http://www.tandfonline.com/doi/abs/10.1080/15265161.2013.868950?journalCode=u
ajb20

Heiby, J. (1998). Quality Assurance and Supervision Systems. *Quality Assurance Brief* 7(1):
1-3. http://www.ncbi.nlm.nih.gov/pubmed/12294098

Heiby, J. (1993). Project Officer's Perspective: Quality Assurance as a Management Tool. *Quality Assurance Brief* 2(1):1-4. http://www.ncbi.nlm.nih.gov/pubmed/12345155

Hurst, S.A. (2014). Simplicity as Progress: Implications for Fairness in Research with Human Participants. *American Journal of Bioethics* 14(2):40-41. http://www.tandfonline.com/doi/abs/10.1080/15265161.2014.872392

Hutton, J.L.; Eccles, M.P. and Grimshaw, J.M. (2008). Ethical Issues in Implementation Research: A Discussion of the Problems in Achieving Informed Consent. *Implementation Science* 3:52. http://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-3-52

Hyder, A.A., Ali, J., Hallez, K., White, T., Sewankambo, N.K., & Kass, N.E. (2015). Exploring Institutional Research Ethics Systems: A Case Study From Uganda. *AJOB Empirical Bioethics* 6:1-14. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4652948/

Hyder, A.A., Rattani, A. Krubiner, C. Bachani, A.M. and Tran, N.T. (2014). Ethical Review of Health Systems Research in Low- and Middle-income Countries: A Conceptual Exploration. *American Journal of Bioethics* 14(2):28-37. http://www.tandfonline.com/doi/abs/10.1080/15265161.2013.868950?journalCode=uajb20

Hyder, A.A., Bachani, A. and Rattani, A. (2012a). Ethics of Health Systems Research - A Scoping Study. Working Paper (available from authors). Geneva: Alliance for Health Policy & Systems Research.

Hyder, A.A., Krubiner, C.B., Bloom, G. and Bhuiya, A. (2012b). Exploring the Ethics of Long-Term Research Engagement with Communities in Low- and Middle-income Countries. *Public Health Ethics* 5(3):252-62. http://phe.oxfordjournals.org/content/5/3/252

Hyder, A.A. and L. Dawson (2005). Defining standard of care in the developing world: The intersection of international research ethics and health systems analysis. *Developing World Bioethics Journal* 5(2):142-52. http://onlinelibrary.wiley.com/doi/10.1111/j.1471-8847.2005.00109.x/abstract

Ijsselmuiden, C.B. and Faden, R.R. (1992). Research and Informed Consent in Africa: Another Look. *New England Journal of Medicine* 326(12):830-33. http://www.nejm.org/doi/full/10.1056/NEJM199203193261212

J-PAL (undated). Introduction to Evaluations. https://www.povertyactionlab.org/research-resources/introduction-evaluations

Lairumbi, G.M., Michael, P., Fitzpatrick, R. and English, M.C. (2011). Ethics in Practice: The State of the Debate on Promoting the Social Value of Global Health Research in Resource Poor Settings Particularly Africa. *BioMed Central Medical Ethics* 12: 22. http://bmcmedethics.biomedcentral.com/articles/10.1186/1472-6939-12-22

Lavery, J.V. (2004). Putting International Research Ethics Guidelines to Work for the Benefit of Developing Countries. *Yale Journal of Health Policy, Law, and Ethics* 4(2):319-36. http://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1087&context=yjhple

Lee, L.M. and Thacker, S.B. (2011). Public Health Surveillance and Knowing about Health in the Context of Growing Sources of Health Data. *American Journal of Preventative*

Medicine 41(6):636-40. http://www.ajpmonline.org/article/S0749-3797(11)00674-X/abstract

Lee, L.M.; Heilig, C.M. and White, A. (2012). Ethical Justification for Conducting Public Health Surveillance without Patient Consent. *American Journal of Public Health* 102(1):38-44. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3490562/

London, A.J. (2000). The Ambiguity and the Exigency: Clarifying "Standard Of Care" Arguments in International Research. *Journal of Medicine and Philosophy* 25(4):379-97. https://www.cmu.edu/dietrich/philosophy/docs/london/Standard_of_Care.pdf

London, A.J., Borasky, D.A. Jr., Bhan, A. and Ethics Working Group of the HIV Prevention Trials Network (2012). Improving Ethical Review of Research Involving Incentives for Health Promotion. *PLoS Medicine* 9(3):e1001193. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001193

Miller, F. G. and H. Brody (2002). What Makes Placebo-controlled Trials Unethical? *American Journal of Bioethics* 2(2):3-9. https://muse.jhu.edu/article/327

Olson, N.W. (2014). Conceptualizing Ancillary Care Obligations in Health Systems Research. *American Journal of Bioethics* 14(2):46-7. http://www.tandfonline.com/doi/abs/10.1080/15265161.2013.868959?journalCode=uajb20

Peters, D.H., El-Saharty, S., Siadat, B., Janovsky, K. and Vujicic, M. (2009). *Improving Health Service Delivery in Developing Countries*. Washington, DC: International Bank for Reconstruction and Development/The World Bank. https://openknowledge.worldbank.org/handle/10986/12335

Pratt, B. (2014). Connecting Health Systems Research Ethics to a Broader Health Equity Agenda. *American Journal of Bioethics* 14(2):1-3. https://minerva-access.unimelb.edu.au/bitstream/handle/11343/56562/AJOB%20editorial-%20Submission.pdf?sequence=1

Protection of Human Subjects Research (2009) *Title 45 Code of Federal Regulations Part 46*, Washington, DC: US Department of Health and Human Services, www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102 (accessed 12 March 2015)

Reinke, W.A. (1995). Quality Management in Managed Care. *Health Care Manager* 2(1):79-88. http://www.ncbi.nlm.nih.gov/pubmed/10165645

Rennie, S. (2014). Tinkering with the Health of the Poor. *American Journal of Bioethics* 14(2): 43-44. http://www.tandfonline.com/doi/abs/10.1080/15265161.2014.871921?journalCode=uajb20

Sim, J. and Dawson, A. (2012). Informed Consent and Cluster-randomized Trials. *American Journal of Public Health* 102(3): 480-85. http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.2011.300389

Swain, G.R., Burns, K.A. and Etkind P. (2008). Preparedness: Medical Ethics Versus Public Health Ethics. *Journal of Public Health Management and Practice* 14(4):354-57. https://www.researchgate.net/publication/5303558_Preparedness_Medical_Ethics_Versus_Public_Health_Ethics

Taljaard, M., Weijer, C., Grimshaw, J.M., Belle Brown, J., Binik, A., Boruch, R., Brehaut, J.C., Chaudhry, S.H., Eccles, M.P., McRae, A., Saginur, R., Zwarenstein, M. and Donner A. (2009). Ethical and Policy Issues in Cluster Randomized Trials: Rationale and Design of a Mixed Methods Research Study. *Trials* 10:61. http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-10-61

Taylor, C.E. (1984) *The Uses of Health Systems Research*, Geneva: World Health Organization, http://whqlibdoc.who.int/php/WHO_PHP_78.pdf

Thompson, A.K., Faith, K. Gibson, J.L. and Upshur R.E.G. (2006). Pandemic Influenza Preparedness: An Ethical Framework to Guide Decision-Making. *BioMed Central Medical Ethics* 7:E12. http://bmcmedethics.biomedcentral.com/articles/10.1186/1472-6939-7-12

Upshur, R.E.G. (2002). Principles for the Justification of Public Health Intervention. *Canadian Journal of Public Health* 93(2):101-03. http://www.jcb.utoronto.ca/people/publications/cjph2002.pdf

Wallwork, E. (2008). Ethical Analysis of Research Partnerships with Communities. *Kennedy Institute of Ethics Journal* 18(1): 57-85. http://www.ncbi.nlm.nih.gov/pubmed/18561578

Weijer, C. and Emanuel, E.J. (2000). Ethics: Protecting Communities in Biomedical Research. *Science* 289.5482: 1142-44. http://med.stanford.edu/content/dam/sm/bioethics/resources-secure/Science2000Weijer.pdf

Weijer, C., Grimshaw, J.M., Taljaard, M., Binik, A., Boruch, R., Brehaut, J.C., Donner, A., Eccles, M.P., Gallo, A., McRae, A.D., Saginur, R. and Zwarenstein M. (2011). Ethical Issues Posed by Cluster Randomized Trials in Health Research. *Trials* 12:100. http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-12-100

World Health Organization (2009). Scaling up Research and Learning for Health Systems: Now is the Time. Report of a High Level Task Force (presented and endorsed at the Global Ministerial Forum on Research for Health 2008), Bamako, Mali, 17−19 November. http://www.who.int/rpc/publications/scaling_up_research/en/

Zeitz, P.S., Salami, C.G., Burnham, G., Goings, S.A., Tijani, K. and Morrow R.H. (1993). Quality Assurance Management Methods Applied to a Local-Level Primary Health Care System in Rural Nigeria. *International Journal of Health Planning and Management* 8(3):235-44.

Chapter 7: Collecting the Evidence

Henry Lucas
Institute of Development Studies

1.   Data collection and resource allocation

In chapter 1 it was argued that where possible implementation researchers should be embedded as members of the implementation team, equally committed to ensuring a successful outcome and fully engaged in decision-making processes. One important area of decision-making concerns the allocation of scarce resources - which could alternatively be used to improve the scope or quality of the intervention - to data collection activities. Following the proposed definitions of chapter 1, these activities can be seen as directed towards three identifiable objectives:

1. Intervention management and accountability;
2. Operations Research intended to improve the current implementation;
3. Implementation Research to learn lessons from the current implementation that can be used in scale-up or re-location to a new context.

However, these objectives can be seen as highly interrelated, each involving the need to track implementation progress against the original intervention design, identify potential weaknesses in that design and test the initial assumptions on which it was based.

Every intervention should have budget lines intended to address objective 1, covering the costs of collecting the data required for intervention management and to demonstrate to those providing funds that resources are being allocated appropriately and outputs produced as intended. For convenience we can regard all such planned expenditures as falling under the general heading of intervention 'monitoring and evaluation'. Often this will include a separate item to meet the cost of operations research studies to be undertaken in pursuit of objective two, for example testing alternative approaches to service delivery or behaviour modification in order to determine which would best serve the needs of the intervention. As discussed earlier, funding of the implementation research activities, including researcher time and data collection costs, will typically not be included in the intervention budget and would often be provided by another agency. However, much of the data required by the implementation researcher can be derived from that collected under the intervention budget, though the analysis of that data may well differ. This provides an opportunity to establish a mutually beneficial arrangement, with implementation researchers providing support, and possibly additional resources, to the intervention monitoring and evaluation system in return for full engagement in the design of that system.

From an implementation research perspective, the design of monitoring and evaluation systems requires not only an understanding as to how the data generated will allow rigorous analysis of the implementation process and the interaction of that process with key contextual factors, but an awareness of the types of evidence that will be acceptable to different stakeholders and audiences (Murray 2007). Apart from the other members of the implementation team, these might include national/ local policy-makers and officials; health workers; NGOs; donor and other international agencies; beneficiary communities and the general population. For example, donor agencies may demand 'objective' quantitative outcome indicators, while communities may be more impressed by qualitative evidence that reflects their own perceptions and concerns. A further consideration is that traditional monitoring and evaluation systems can tend to follow a routine reporting and analysis plan which is insufficiently responsive to rapidly developing potential opportunities and threats to

the implementation process, especially in a CAS context. Both operations research and implementation research activities may often benefit from ad-hoc, 'real-time' exercises, possibly undertaken in collaboration with service providers or intended beneficiaries, which can be effective in providing rapid feedback on access barriers and process bottlenecks as they arise. Such activities can combine information gathering with exercises that explore ways in which these barriers or bottlenecks may be overcome, both in the existing or in future implementations.

Choice of research methods

There is a tendency for researchers to think in terms of undertaking studies that can be easily categorized, using labels such as quantitative, qualitative, participatory, action-research, desk research, etc. A better approach is to start from a careful review of the various research questions that need to be addressed and then to assess which research methods might be able to deliver the required information on each of these questions. Resources should then be allocated in such a way as to best meet the overall research objectives, which will usually involve prioritising some questions over others, within whatever 'budget constraints' apply - which may relate not only to financial limitations but also to the limited availability of time, skilled/experienced personnel, access to data sources, etc. Such an approach may often involve unpalatable compromises relating to the scope, depth or precision of the intended research findings, but setting unattainable goals or attempting to 'make do' with inadequate resources will almost certainly degrade the *quality* of those findings.

One complicating factor in adopting such an approach, is that most researchers have a strong preference for primary data collection. They identify what they see as their requirements for specific data items and then assume that those requirements can only be met by the careful design and application of data collection instruments that are intended to deliver those items. However, before deciding to invest in any substantial data collection exercise, which will almost inevitably be costly and will typically prove substantially more costly than anticipated, it is almost always worthwhile to undertake a systematic inventory of relevant, accessible, secondary sources. These will almost certainly not provide precisely the data you *want* but may well provide data that can meet at least some of your underlying *needs*.

For example, reports and/or data from previous income or expenditure surveys may provide a reasonably adequate guide to current distributional questions if there is no reason to suspect that these may have changed radically since those surveys were undertaken. Even poorly maintained hospital financial records may provide better data for the estimation of inpatient treatment costs for a given condition than can be obtained from a survey that relies on the memories of former patients. Careful study of official reports, even if you are doubtful as to their reliability, will often enable you to be much more efficient in undertaking key informant interviews with senior policy-makers, allowing you to focus on questions which test the veracity of the information and opinions in those documents.

The general proposition here is that *all* potential sources of relevant data should be explored and their availability, accessibility, cost and potential value assessed before deciding on your research strategy. These would include:

- Documents: official reports, academic journals, media articles, internet blogs, etc.
- Routine data systems (RDS): financial data, personnel data, clinic records, etc.
- Existing survey data: national surveys, Demographic and Health Surveys (DHS) World Health Surveys (WHS), Multiple Indicator Cluster Surveys (MICS), etc.
- Implementation RDS: from the implementation monitoring system.
- New sample surveys: of patients, facilities, providers, community members, etc.
- Qualitative studies
- Rapid appraisal and/or Participatory exercises.

2.   Secondary Data

Document Review

Document review should involve the *systematic* compilation and analysis of relevant printed and electronic material. In terms of health systems interventions, probably the most important sources will be legislative documents, policy statements and which set out the basic frameworks within which health systems function. There will also be a wide range of regulations, guidelines, manuals, protocols, etc., issued by ministries, other official agencies or by facilities themselves, which define the detailed operational procedures that should be followed in the management, administration and delivery of health services. These documents can be important even if the researcher is fully aware that they are widely disregarded, in that they can indicate what individuals perceive to be appropriate behaviour in terms of health service provision or at least what they perceive as being acceptable to the general population.

Organisations and individuals will often try to behave as set out in such documents even in the most difficult and chaotic circumstances, following procedures they know to be irrelevant simply because they have no well-defined alternative mode of operation. Working in Nigeria in the mid-1990s, when public health services were almost non-existent in many rural areas, the author had to work around a legal prohibition on the use of alternative forms for the collection of data on public facilities. This was often cited by providers even though it was clear to all those concerned that the official health information system had ceased to function. Similarly, state government officials would expend considerable efforts on the careful preparation of annual budgets, even though they knew that these would have limited effect in terms of controlling actual expenditures. Analysis of the gulf between what is contained in such official documents and the reality on the ground is often key to understanding the context within which interventions are undertaken.

A systematic document review should aim to at least consider, if not analyse, all those materials which may be relevant, from whatever sources. This will be time consuming and should not be seen as an activity which takes place only at the start of the implementation process but one which can be conducted at a steady pace over the research period. Increasing use of media outlets and in particular of the internet has dramatically increased the volume of information that is relatively easily accessible to the researcher. For example, reports from the international Demographic and Health Surveys, World Health Surveys and Multiple Indicator Cluster Surveys indicated above can be inspected or downloaded from their websites. In many countries census and survey reports are often made available in reasonably timely fashion through the internet sites of national statistical agencies or ministries. Reports from earlier periods, possibly useful in considering trends over time, may also have been archived on the International Household Survey Network (HISN) website. Expectations should be limited. Survey reports tend to provide relatively simple statistics, often at a high degree of aggregation, using variables that will almost certainly not have been defined as you might wish. Nonetheless, they can often provide a limited number of apparently relatively reliable indicators which may be very useful in terms of confirming or challenging information received from other sources.

Interesting insights into the concerns and intentions of relevant organisations can often be gained by examining their press releases, which again are now often made available via the internet. Given that they are almost always intended to present the organisation in a favourable light, these need to be subjected to careful analysis and interpretation, but can be extremely useful in determining the most effective strategy for exploring their underlying aims and objectives, for example in the design of key informant interviews. Media articles - in newspapers/magazines or on television or radio - provide another relevant source, which in this case will need to be assessed in terms of an informed judgement as to whether the author

can be seen as independent or biased in one direction or another. Such biases do not render the information useless - as long as its implications can fully incorporated into your analysis. Articles based on the opinions of those critical of your intervention can be of particular interest in terms of understanding the arguments that an implementation may have to address and in revising your stakeholder analysis.

A related and underutilised source are the internet 'blogs' that may be written by individuals within a healthcare agency or community based organisation, either on their own websites or on social media sites such as Facebook. The author gained valuable insights into the problems faced by an agency concerned with providing health advice from the activity on one such site, where it is easy for contributors to forget that their discussions are open to public view. Other sources of interest include the many advertisements for health providers and products, which may play a major role in influencing the attitudes of the local population as to the availability of treatments for a range of conditions. These may be found in the local media or on the internet but are also widely displayed on posters, either positioned by roads or in shop windows. Such 'documents' can now easily be captured and analysed in the same way as other materials using digital cameras. Such cameras can also be used to incorporate a range of relevant maps, charts and photographs which the researcher may encounter.

It should be emphasised that document review is a research activity and as such should be fully described in the research report. Details must be provided as to how different sources were explored and relevant materials identified, accessed and analysed. Stage one in such an analysis, as indicated above, is to understand the origins of the document and the reasons for its production. This should allow you to make an intelligent assessment as to how it may be interpreted. Was it an uncontentious attempt to codify existing practices to make sure that all providers followed a common approach or a highly contested regulation which imposed unwelcome constraints on income generating activities? Was it intended to demonstrate how careful a government agency had been in managing a social insurance fund or to attack the profligacy of political opponents? Stage two would involve seeking ways to verify any of the claims or estimates contained in the document. Can they be compared with those from any other source? Is the methodology adopted described in the document and if so does it seem appropriate? Is it possible to discuss the findings with those who had published them? Stage three involves the more difficult tasks of extracting relevant excepts from each document summarising these without losing essential content and then combining these summaries under various themes and sub-themes. This is most frequently undertaken on a relatively pragmatic basis, relying on the experience and skills of the researchers. However, there are more rigorous methodological approaches which are usual described under the heading of 'content analysis' (Hsieh and Shannon 2005) and in recent times these have often been undertaken using a range of specialised computer software packages (ESRC undated).

Routine Data Systems

Routine (administrative) data systems (RDS) have the great potential advantage that they can deliver disaggregated, time-series data by geographical area (region, state, district, sub-district, etc.) (Lagarde 2012). Facility records, for example attendance registers, patient records, disease registers, prescriptions, insurance payments, financial accounts, etc. can be an important source of quantitative data, if they can be accessed by the researcher. They may be immediately of value or of use after further processing. For example they may require reorganisation, aggregation, disaggregation or other manipulation. In this case it is necessary to ensure that the nature of the data in terms of such aspects as definitions and collection procedures is thoroughly understood, as the possibilities for misinterpretation are considerable. It may in some cases be cost-effective to invest resources in measures which support improvement of the RDS. For example, in some countries many primary facilities still lack simple electronic calculators and may have to spend considerable time adding up many columns of figures, often making mistakes. In one study the author found that the simple provision of higher quality attendance registers and prescription pads (the only documents

available at this level) and a supply of pens, pencils and erasers, dramatically improved the recording of patient and treatment details over the course of the research. In recent years the provision of mobile phones does appear to have considerable potential (Neupane et al. 2014, Uganda 2012).

Mainly because they can provide data which can be disaggregated to the particular location in which an intervention is undertaken, it can be more useful to analyse RDS than existing survey data. However, in most countries it suffer from well-known limitations, often in spite of many attempts at improvement. There are three major issues:

- coverage - focusing only on those who use services can be extremely misleading - we know that it is generally the poor and vulnerable who are most often excluded;
- general poor quality (accuracy, timeliness) - often reflecting indifference on the part of staff who have come to believe that senior health service managers, officials and politicians rarely make use of, or even consult, the data they provide;
- incentives to misreport (e.g. where providers receive performance related payments) - and an absence of effective audit systems that might detect misreporting.

The poor quality of the data may be improved to a limited extent by measures such as those indicated above, but typically relate to a widespread culture of indifference to reliable reporting which is not easily amenable to change, given the resources available to any specific intervention. Some financial data (e.g. payroll data, payments by health insurance agencies) may be more reliable because it is subject to audit procedures. One potentially useful activity is to explore the possibilities for combining routine data with other sources, such as surveys, to generate 'best estimates' (Rowe 2009). This implies the need for expectations to be limited and second-best options to be explored. For example, while such basic indicators as service utilisation, access and cost are not ideal, they may provide a reasonable basis for context analysis and to verify data from other sources.

RDS data quality is a particular concern when disaggregation within the intervention area is required. As a general rule, administrative data quality depends on the quality of administrators, and both tend to be correlated with the overall level of development. The poorest areas and facilities typically have the least reliable data. This is of particular concern in terms of indicators derived from information systems which are subject to the pressures associated with the provision of marketable goods and services. For example, rural health workers in poorer areas (given that their government salaries are sometimes barely sufficient to purchase basic food and clothing) have become very adept at providing information that satisfies higher levels of administration while not limiting their alternative income generating activities. It should be noted that variations in the quality of data, particularly administrative data, between areas and facilities may also influence aggregate estimates, as these are often based on partial coverage. Facilities in less developed areas not only tend to provide less reliable data, they often fail to provide data on time. As overall estimates are often derived by "grossing up" the information available when estimates are required - i.e. information from better resourced facilities, biases which tend to overestimate service utilisation, staffing levels, drug availability, etc. may be introduced.

Finally, note that many of the most important health indicators require the combination of service data from the RDS with overall or age-specific population estimates. These will reflect the 'denominator problem' of indicator construction - the fact that these estimates are typically crude estimates and/or outdated. The influence of changing population sizes and distributions, often due to internal migration, on access and utilisation measures can be substantial and will often need to be considered in the interpretation of trends over time. Again, poor regions may be particularly affected by both push and pull migration factors. The use of population estimates also raises issues of data availability. Population estimates in years removed from that in which the census is taken will be derived from demographic models, often based on

parameters estimated from DHS data. These models may be reasonably reliable at the national level but are not intended for sub-national estimation and typically do not allow for the effects of possibly large-scale internal migration.

Existing Survey Data

Anyone who has undertaken a reasonably large-scale sample survey will appreciate that it can be a daunting task. It primarily requires a range of managerial and administrative skills that are often lacking even in some of the most talented and experienced social science researchers. In particular, surveys usually involves the hiring, training and management of a substantial number of enumerators, supervisors and data entry staff who may have little interest in the survey objectives and need constant encouragement and oversight to ensure the quality of the data produced. They also typically involve a considerable investment in terms of both time and money. If it seems at all possible that relevant research questions can be addressed by secondary analysis of an existing survey data set which is known to be of reasonable quality, it would be a mistake not to at least serious consider this option (Boslaugh 2007).

One key question to be addressed is how the quality of the existing data set is likely to compare to that from any new survey. Where surveys which have been conducted on a regular basis for a number of years by permanently employed staff members, for example from a national statistics office, their accumulated experience may well imply that the quality of the final product is likely to be considerably in advance of that from a newly designed survey conducted by a team of recent recruits employed on short term contracts. The sampling expertise and resources (e.g. computerised sampling frames) available within the agency that designed the existing survey may also have been far in advance of that available within the implementation research team. This may imply that there will be greater uncertainty as to the validity of estimating population parameters using sample statistics derived from the new survey.

In addition, the existing survey may have included questions on topics, for example incomes or expenditures, which would be of considerable value in any analysis but which could not realistically be included in a new survey given budget constraints. If it had been undertaken on a national or sub-regional basis, it could also provide an opportunity for direct comparison of data from the implementation sites with that from other areas, an important consideration when exploring the opportunities and challenges involved in scaling-up or relocating the intervention. Similarly, if the same questions have been asked in successive rounds of the survey over previous years, it may allow analysis of trends over time which provide insights that would not be available from a cross-sectional survey.

The above qualities are of course irrelevant if the survey data cannot be used to explore the questions that the research needs to address. An initial problem may be that it is difficult simply to gain access. For example, national statistical agencies will usually argue that survey data is collected on the basis that it will only be used for a specific purpose and that the respondents have been assured that it will not be shared with other organisations. Versions of the data from which any variables that can be used to identify respondents have been removed may be made available but often only with a considerable delay that reduces its value. Agencies may also require researchers to make a formal request for the data which involves a detailed explanation of the types of analysis to be performed and the intended uses of any findings, possibly requiring any resulting reports to be submitted to them before dissemination. In some cases they may also demand a substantial payment for use of the data. Note that the international agencies indicated above usually do make data freely available to researchers with minimal formality and it is also worthwhile to explore the International Household Survey Network (IHSN) website, which does hold selected survey data sets, though many of these will be some years out of date.

The researcher will not only need to gain access to the data itself but also to the 'meta-data' which provides a detailed description as to how it may be analysed and interpreted. As a minimum this must include the questionnaires and coding manuals, but it will often be very useful also to have copies of the enumerator and supervisor manuals. For example, if respondents were asked if they visited a public or private clinic the answers may well differ substantially depending on the guidance (if any) provided by enumerators as to how to distinguish between these two types of facility. Having considered the precise nature of any variables of potential interest in the data, the researcher will then have to make a considered decision as to whether they can be used to at least provide insights into the original research questions. It is in the nature of secondary data analysis that the variables available are rarely those which the researcher would have chosen to analyse. The original question may not have been worded as you would have wished. The instructions to the enumerators may have resulted in an excessive number of missing responses. The coding system adopted may have lost information that would have been extremely useful. Even if these problems can be overcome, you may find that the sample size is too small to allow disaggregation to the extent necessary to provide relevant estimates for the intervention population. Nonetheless, all these potential limitations should not prevent you from exploring this option. The costs are often minimal and the potential benefits considerable.

3. Primary Data Collection

Qualitative or Quantitative?

A somewhat simplistic view of the appropriate uses of alternative approaches to primary data collection is shown in the following table:

Table 1: Matching methodologies to objectives

Objective	Methodology
Quantitative estimates representative of population parameters Knowledge of sampling errors	Formal surveys
Quantitative data with some understanding of processes Repeatable for trend assessment	Quantitative Rapid Appraisal
In-depth knowledge of behaviours, perceptions, attitudes, etc. Interpretation of existing quantitative data	Qualitative Methods
Very limited contextual knowledge Limited resources	Qualitative Rapid Appraisal

At one extreme, we might be concerned to produce estimates of specific population parameters, for example utilisation rates for health facilities or the frequency of given symptoms in young children over a previous period, together with the associated 'errors of estimation', which allow us to specify how confident we can be that those estimates fall within a given range. If we wish these estimates and confidence limits to be widely accepted as valid, we would be well advised to use formal surveys which follow the accepted principles of statistical inference. If we are less concerned about the precision of such estimates and believe that we can derive them to an acceptable degree of accuracy using alternative and less resource intensive methods, for example by extrapolation from facility records, questioning key informants or focus groups, or techniques such as participatory ranking or mapping (Chambers 2007, Rifkin 1996). In the above, such approaches are described as 'quantitative rapid appraisal'. Using standard procedures can allow comparison between areas

and over time, but the extent to which such estimates are accepted will in this case depend on our ability to persuade others of their reliability.

If we need to understand not only *how* individuals and organisations behave but *why*, we may decide that some form of detailed qualitative study is required. This may involve long-term engagement with the study population, using a range of observational and interview techniques to formulate and then test alternative explanatory theories. Finally, if we know very little about the context within which we are working, a common situation at the start of any research activity, we might adopt an approach which can for convenience be labelled 'qualitative rapid appraisal', mainly using key informant interviews to enable us to at least frame relevant research questions.

However, it is often worthwhile to think 'outside-the box' when considering which methodologies and methods might be the most appropriate (or cost-effective) to meet data requirements in a specific context (Holland 2007). Kanbur (2003) suggests that we usually categorise qualitative and quantitative methods as having the following characteristics, locating them at the opposite ends of five 'dimensions':

Qualitative		Quantitative
Non-numerical information	➡	Numerical information
Specific and narrow target groups	➡	Large general target population
Active engagement with respondents	➡	Passive involvement of target population
Inductive methods of inference	➡	Deductive/statistical inference
Description/generalisation/theory construction	➡	Hypothesis testing/econometric modelling

But there are no 'rules' that force you to accept this dichotomy. Given that researchers are always constrained by limited budgets, they should try to assess the costs and benefits of locating at different points along each of these dimensions in a specific research context, considering only how they will justify their decisions if challenged. For example:

- Traditional household surveys can be used to gather non-numerical information using 'open-ended' questions (Rog et al, 2011).
- Participatory methods can be used to generate numerical data - e.g. ranking of providers, estimated travel times to different facilities, etc. (Chambers, 2007).
- Qualitative studies can use probability sampling and large sample sizes to gain credibility (Barahona and Levy 2002).
- Qualitative studies may rely primarily on observational data, involving limited interaction with members of the targeted population (Walshe et al. 2012).
- Qualitative studies of social networks can use statistical methods and mathematical modelling techniques to generate network maps (Bishai et al. 2014).

Potential advantages and disadvantages of qualitative studies
One great attraction of qualitative approaches to many researchers is the extent to which they feel in control of the process. Sample sizes are typically relatively limited, allowing a small number of skilled, experienced researchers to take the time required to fully engage with those who are providing information. There can be considerable flexibility, with those researchers being trusted to make decisions as the research proceeds, for example selecting additional or alternative respondents, adapting questions or participatory as their knowledge of a situation increases and possibly opening up unplanned lines of enquiry if unexpected responses or observations suggest that these may be of importance.

Given sufficient expertise, researchers can undertake detailed investigations not only as to the knowledge of respondents but also their perceptions, attitudes and motivations. If they can gain their trust, they may be able to explore sensitive issues and assess emotional responses. Interviews which take place in homes or facilities will often allow valuable insights into relationships, processes and contexts simply by careful and prolonged observations. Of particular importance when there is limited knowledge at the start of a research activity as to which are the most relevant issues, qualitative studies can allow the gradual elaboration of concepts and theories as the research proceeds, delaying the often very difficult task of formulating precise definitions of variables and the expected relationships between them until the researcher has had an opportunity to experience the ground realities (Kuznetsov et al. 2013).

To some extent, the disadvantages associated with the archetypal qualitative study can be seen as the mirror image of the advantages. The flexibility which is so attractive to many researchers tends to place great weight on the regard in which the members of the research team are held by those whom they might wish to persuade of the value of their findings. The central issue is that of subjectivity, that given the extent of their control over the process of data collection it is likely that the research findings will be at least partly determined by the preconceptions of the researchers, i.e. they will tend, quite possibly unconsciously, to gather information that reinforces their personal perceptions as to how the world works. While it can reasonably be argued that quantitative research also has to contend with this issue, the use of predetermined instruments and procedures - questionnaires, manuals, sampling designs etc. - provides those who wish to determine the extent to which findings have been influenced by the decisions of the researchers with the documentary evidence they require. This indicates the way in which qualitative researchers can guard against their findings being dismissed as 'too subjective', by ensuring that every step in the data collection process is carefully documented, providing detailed descriptions not only of what was done but why. This should be a central component of an activity usually described as 'reflexivity' (Finlay 2002; Mruck and Breuer 2003), ongoing assessments by each researcher of the extent to which their activities might be driven by personal factors and attempts to counteract that tendency.

One related common criticism of qualitative studies is that of sample selection biases, for example tending to gather information more from those who are in favour or those who are against the intervention, neither group being *representative* of the overall population. Researchers will usually try to avoid obvious potential biases, such as relying on local officials or 'community leaders' to determine their subjects, but it is easy to overlook other potential pitfalls, for example limited resources may result in a failure to seek out less accessible stakeholders, for example those who live in remote or less accessible areas. Sample sizes are often limited in qualitative studies. The essential need to use only capable, experienced researchers, because the quality of the findings is so dependent on their abilities, generally implies that the cost per respondent will be substantially higher than that for quantitative studies using enumerators to complete standardised questionnaires.

Small samples can raise difficult problems in terms of analysis and interpretation, given that we are often interested in the relationship between the diverse circumstances and characteristics of our respondents and their perceptions, attitudes, etc. We would often see it as essential to distinguish between respondents in terms of a range of attributes including gender, age-group, income/wealth, rural/urban, etc. even if we adopt a policy of stratification, such that we have respondents in each cell of the implied multi-way table, the numbers in each cell will be so small that we may be reluctant to infer that they can be extrapolated to other 'similar' individuals in the study population. One common challenge to qualitative findings is that they are *anecdotal*, interesting as descriptions of individual cases but *unrepresentative* and therefore of limited use in terms of reaching general conclusions and hence in terms of policy making. A similar complaint may arise with respect to comparisons between the various

groups, for example differences in attitudes as between men and women. If there were relevant differences in the nature of the information gathering process between groups, for example different researchers choosing to vary the type or sequence of questions, or the use of male researchers to interview men and females to interview women, it might be argued that at least part of the observed differences may simply reflect inter-interviewer variation. A final, practical disadvantage, of qualitative studies is the sheer volume of information, mainly textual, that they almost always generate, posing substantial problems in terms of analysis and interpretation, even with the use of computer software packages (ESRC undated).

Potential advantages and disadvantages of quantitative studies

A well designed and implemented probability sample survey has the unique advantage of being able to provide reliable, bounded estimates of key population parameters, for example immunisation rates, illness prevalence rates, utilization of services, average length of stay in hospital, median cost of an outpatient visit, etc. Unlike any other methodology, it allows the researcher not only to generate such estimates but to specify how 'confident' they are that each estimate falls within a stated range (the 'precision' of the estimation). These estimates are derived using the area of mathematics known as statistical inference, which allows a researcher who can show that they have 'followed the rules' of probability surveys to present such estimates without the need for further justification. While other approaches, for example market research surveys or political opinion polls, may make similar claims, they are almost always not following the rules and therefore cannot legitimately use the language of statistical inference to support those claims.

This ability to generate reliable estimates to a given level of precision can be very attractive to policy-makers because it allows them to assess the potential quantitative impact of a given intervention. For example, China has recently started to introduce policies which provide improved health insurance coverage for the poorest members of rural populations. Such policies had been recommended by health researchers for many years but became much more acceptable to government when the costs of such changes could be reliably estimated from probability sample survey data. Again very useful from a policy perspective, the adoption of predetermined and standardised instruments for data collection in most quantitative studies enhances the credibility of making comparisons between different subgroups of the target population. Given that precisely the same questions are asked in what should be precisely the same manner to such subgroups, for example the heads of richer and poorer households, it will often seem plausible to directly compare their responses, for example in terms of the proportion of children under two vaccinated against polio. Quantitative studies generally try to minimise any variations in behaviour between those collecting data from different subgroups, which may be misinterpret as between sub-group variation. Similar consideration apply to comparisons over time, for example estimation of trends in childhood malnutrition rates using DHS data for different years.

The desire to make comparisons between subgroups or over time is related to one of the main disadvantages of the typical quantitative approach, the difficulty of developing simple, *uniformly applicable* definitions of key concepts that are well understood and have a common interpretation across all subgroups of the population. For example, in one pilot exercise conducted by the author, a standard question as to whether anyone in a household had suffered an acute sickness in the previous two weeks produced incidence rates for those in the poorest rural area surveyed that were far too low to be believable. A follow up qualitative study found that fevers were so common that may people did not consider them worth reporting. Similar issue arise with respect to many of the covariates on which we often try to collect data in such surveys. The distinction between rural and urban areas, for example, is often problematic as is that between public, not-for-profit and for-profit facilities (if we are aware that they are all charging for services to a greater or lesser extent). Particular difficulties arise with studies that are concerned with equity. Measures of income, expenditure, wealth,

indebtedness, vulnerability, etc. are notoriously difficult to define in ways which can be confidently expected to produce comparable findings across sub-groups (World Bank 2003).

The above indicates the need for a profound understanding of both the topics addressed and the population targeted at the design stage of any quantitative study. They should certainly not be used to explore issues about which the researchers have very limited understanding. That will almost always result in a substantial expenditure of resources to little purpose. In-depth knowledge is essential if the study is to be well designed and the design phase is often the key to a successful outcome. The implementation of a large-scale quantitative study is primarily an exercise in human resource management and logistics. Once launched it is very difficult to change course or rectify any major design defects that may become apparent. It is essential to ensure: (a) that the research team has the necessary management skills required and that those with these skills are willing to take a leadership role - along with the responsibility for ensuring that the exercise proceeds with as little divergence from the original intention as possible; and (b) that the resources are sufficient to allow for unexpected problems - bureaucratic delays, equipment failures, illness, bad weather, etc. - which will almost inevitably be encountered. Attempting to stretch an inadequate budget and 'hoping for the best' is a recipe for failure. Finally, it should be taken into account that those who most strongly favour quantitative studies often have a tendency to pay insufficient attention to likely data quality issues, preferring to make heroic assumptions as the reliability of the findings derived from this data, often substituting technical expertise for considered analysis and claiming general validity for what are typically very simplistic models of causality.

Combined Methods

In practice, it would be very unusual, and almost certainly a mistake, not to use both quantitative and qualitative approaches in any implementation research exercise. While there is a very long history of researchers combining quantitative and qualitative methodologies, the mid-1990s saw a more formal discussion of the opportunities and potential pitfalls of using 'combined methods' (sometimes described as mixed methods or qual/quant) (Qualquant 2016, Palinkas et al. 2011, Barnett et al. 2016). Most attention has focused on the potential advantages of using qualitative studies to complement and support large scale surveys (Kanbur and Shaffer 2005) they include:

- The use of qualitative studies to improve survey design
- The interpretation of counterintuitive or surprising findings from surveys
- Explaining the reasons behind observed survey outcomes
- Exploring the motivations underlying observed behaviour
- Suggesting the direction of causality
- Assessing the validity of quantitative results
- Understanding conceptual categories such as ill-health, household, etc.
- Interpreting local categories of social differentiation, e.g. poor/non-poor
- Provide a dynamic dimension to cross-sectional household survey data.

However, there are multiple pathways by which qualitative and quantitative studies might be linked. Marsland et al. (1998) categories these pathways under three broad headings:

A: Swapping tools and attitudes: "Merging"
1. Adopting standard sampling techniques in qualitative studies (Barahona and Levy, 2002).
2. Coding responses to open-ended questions using qualitative enquiries.
3. Using statistical techniques to analyse quantitative data obtained from qualitative studies, for example:
   a. Creating frequency tables from coded responses to open-ended questions.

     b. Constructing models based on binary and categorical data from ranking and scoring exercises.
4. Using participatory mapping to create sampling frames for questionnaire surveys.
5. Using findings from qualitative studies to reduce the non-sampling error (e.g. misunderstandings, offensive questions) in questionnaire surveys.

B: "Sequencing"
1. Using exploratory techniques to establish hypotheses which can be tested through questionnaire surveys.
2. Using a questionnaire survey to gather responses to a few key questions from a probability sample of respondents and then undertaking a qualitative follow-up study of respondents that appear to be of particular interest.

C: "Concurrent use" of tools and methods from the different traditions
1. Using a questionnaire survey to determine quantitative indicators (e.g. Likert scales (Sullivan and Artino 2013)) on perceptions and attitudes relating to public and private health services.

alongside

2. Qualitative exercises (key informant interviews, focus group discussions, participatory exercises) to address the same issues with the aim of gaining greater understanding.

These possibilities are reflected in figure 1 below.

Figure 1: Possible combined methods pathways¶
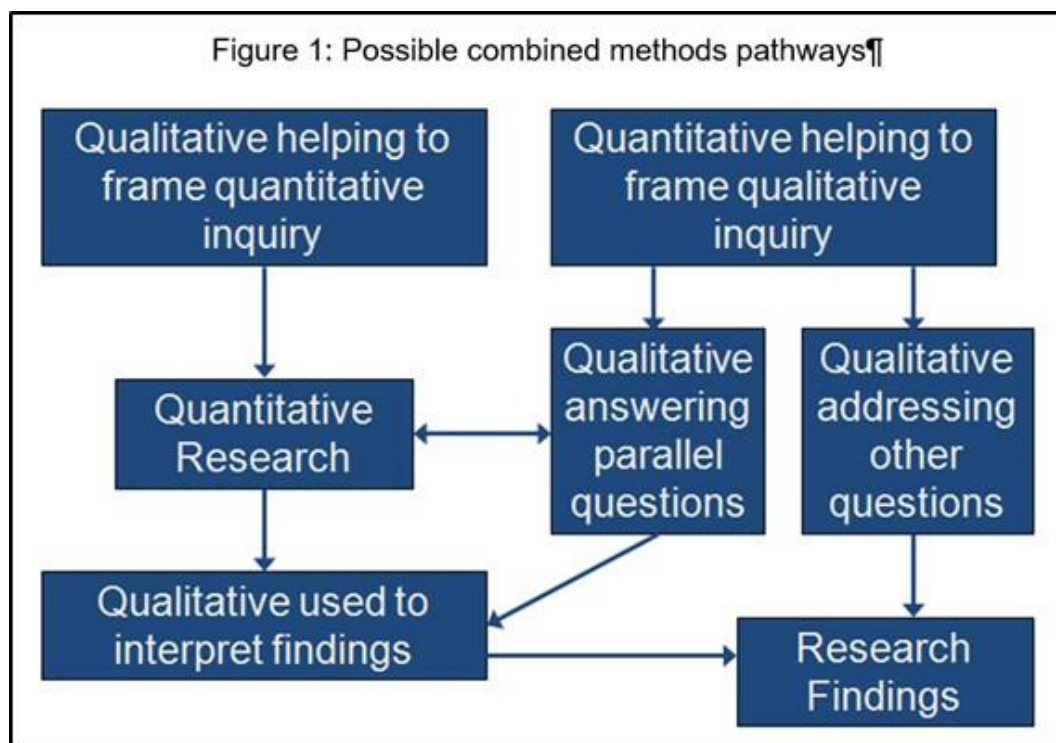
For example, Lucas et al. (2009), used large-scale sample surveys in Cambodia, China and PDR Laos to identify households where at least one member had suffered from a serious illness over the course of the previous year. A limited number of geographical case studies, based on purposively selected counties in China and health districts in Cambodia and Lao

PDR were undertaken. In each of these areas households affected by major illness were identified and studied using a two stage approach:

1. A rapid and reasonably large-scale household questionnaire survey was undertaken using cluster sampling of households within the selected study areas. This aimed to identify households substantially affected by different categories of serious health problems and to estimate the proportions of such households in the population.
2. The sampled households were analysed and classified into a number of strata based on the information provided by the questionnaire survey (the choice of stratification variables is indicated below). In-depth studies, typically requiring 1-2 person days, of a probability sample of the households in purposively selected strata were then undertaken by a team of social scientists.

4.   Sampling in quantitative and qualitative studies

Implementation research aims to generate findings that are taken seriously not only by academics but much more importantly by the implementation team, policy makers, service providers and the general population. As indicated above, it will typically use a combination of quantitative and qualitative methods. However, whatever the methodology there may be advantages in selecting samples that can be seen as representative of some specified population, allowing findings to be more plausibly generalized, or alternatively selecting samples which exhibit a high degree of diversity, to demonstrate that all aspects of an issue are explored. Whatever the methods of data collection, it is at least worth taking time to consider how accusations of bias in the selection of research subjects, a very common tactic adopted by those who wish to discredit unpalatable research findings, might be addressed (Barahona & Levy 2002). A key point to remember is that your choice of data collection procedures does not dictate your choice of sample selection procedures. The only requirement is that these choices should jointly aim to deliver findings of a quality and nature that you believe: (a) justify their use to influence your target audiences; and (b) have the potential to be accepted by those audiences.

Why is sampling important?
Though we may tend to associate the concept of sampling with the formal procedures adopted, for example, in household surveys, quality testing of drug supplies, or audits of financial accounts, in practice it is central to our understanding of the world. Of necessity, almost all of our experiences are derived from samples: the people we have met, the organisations we have encountered, the documents we have read, the places we have visited, are a vanishingly small proportion of all those which might have influenced our perceptions and attitudes. In principle we know that information obtained from samples can be misleading, that it can be *biased*, to a greater or lesser extent, but it is often very difficult to act on that knowledge in a consistent manner. For example, even when engaged in our professional research activities, our impressions of people and places will often be strongly influenced by the small number of individuals that we meet and the observations that we make during a relatively short visit at a particular time of year. Even if we know or suspect that the doctors, nurses, local officials and community leaders that we encounter have been selected by stakeholders that have a strong incentive to ensure that we leave with a particular impression, it is often very easy to persuade ourselves that we are such skilful and experienced researchers that we can 'read between the lines' and make unbiased assessments of the true situation.

There are two problems with this approach. First, that we are almost certainly overestimating our abilities. Self-deception is a widespread human trait even among the most intelligent (Lamba and Nityananda 2014). Second, we are taking a serious risk in assuming that the target audiences for our research findings will accept our assurances that we have such

abilities. As emphasised throughout this book, implementation research findings may be used to advocate major changes in health policy or major reallocations of health resources. It is appropriate therefore for those presented with such findings, particularly if they are in a position to make those changes or reallocations, to aggressively examine the methods we have employed. The simple assertion that you are convinced that the findings are not influenced by *sampling biases* is not, and should not be, a sufficient response.

Some definitions

The language used in discussing sampling procedures can be a cause of confusion. The following definitions can be applied to all forms of sampling:

- Population: A collection of entities - individuals, households, records, organisations - about which we wish to make qualitative or quantitative statements.
- Sample: The subset of entities on which we base those statements.
- Sampling Design: The procedure used to select that subset of entities.
- Sampling errors: Misleading findings arising from reliance on data from a sample.
- Non-sampling errors: Misleading findings arising for any other reason.

Note that some care is required in distinguishing between the sample we intend to obtain and the sample we do obtain. For example, if we make the elementary error of visiting a village at a time of day when most working-age adults are absent, we should be very cautious in assuming that the perceptions and attitudes of those available for interview reflect those of the community at large. Similarly, if any of the questions in a given study address sensitive issues, we may well find that a substantial proportion of the members of our intended sample refuse to answer. Again, we would be foolish to assume that the responses can be considered representative of the overall population. One useful way to think about such issues is to decide which members of the population had at least some chance of being included for our sample. In technical terms, we usually describe such individuals as having a *probability greater than zero* of being included. This group is sometimes described as the *sampling population* because it is the population from which our sample is selected in practice rather than in theory.

We should always be extremely cautious in making inferences from a sample that relate to individuals who were not in the sampling population. For example, an utilisation study based on a sample of clinic returns available at the ministry of health will not necessarily provide reliable estimates at a national level if a substantial number of clinics in poorer, more remote, locations have failed to submit their returns. A study on healthcare costs using a sample of health insurance records cannot be assumed to allow inferences to a population that includes the uninsured. A study assessing satisfaction with health services by using mobile phone calls to a sample of patients following treatment must consider the possibility that the findings might have been very different if it had included those without access to a mobile phone.

Chambers (2006:28-32) uses the term 'rural development tourism' to explore the potential for sampling biases in assessments based on the type of short, infrequent visits to study sites that are common not only in research studies but which also play a major role in the routine activities of public officials, politicians, consultants, NGO/donor agency staff, etc. Such biases, which can dramatically limit the size and diversity of the sampling population include:

1. Spatial Bias: Staying close to urban areas, traveling only on tarmac roads and preferring to visit communities and individuals who live near to such roads.
2. Project Bias: Areas where activities are taking place, projects are in progress.
3. Person Bias: Elites, Males, Users (those who use services or are targeted by projects), Visible/Active Individuals, etc.
4. Seasonal Bias: Fieldwork done when travel is easy, avoiding floods, droughts, etc.

5. Diplomatic Bias: Reluctance to annoy elite members of the community by addressing sensitive issues, tendency to avoid potentially embarrassing encounters with the very poor, the very sick or community outsiders.
6. Professional Bias: Focusing on areas offering favourable opportunities for research. For example, where health facility records are well maintained.
7. Security Bias: Avoidance of areas with a risk of political unrest or violence.

There are many ways in which we can reduce the likelihood and/or extent of sampling bias. For example, in terms of seeking evidence from existing documents we often adopt a process called *systematic review* (Hemingway and Brereton 2009), which aims to ensure (a) that we clearly describe our approach to making best use of those documents and (b) that we have read at least what we can argue (on the basis of explicit criteria) to be the most important. In the situation described above, making a brief visit to a new study site, we might seek alternative ways of selecting our informants, for example asking to talk to male and female, older and younger community members or taking advantage of clinic visits to engage discretely with low level health workers who were not invited to the formal meetings. There are a wide variety of ways to select a sample. The basic principle is that we try to devise *and rigorously implement* sampling procedures that are less likely to result in samples that lead us to make incorrect inferences.

> Example: a small-scale exploratory study in the Niger Delta, Nigeria, sought to discover barriers to use of bed nets (Galvin et al. 2011). Convenience samples of volunteers in a number of villages were asked about sleeping arrangements, perceptions of bed nets and barriers against their use. Because they tended to talk to the most easily accessible respondents, a substantial majority of those in the samples were adult males. The article reports that the "gender bias of our sample" limited "the conclusions that can be drawn … importance of structural and inconvenience factors, safety, and comfort may reflect issues of more concern to males". A "more focused purposive sample" of "mothers of under-fives, and young people 12 to 19 years may illuminate some further barriers or motivations".

Some of the most commonly used sampling methods by researchers, which vary considerably in terms of the extent to which they address the key issue of selection bias, depending on the context in which they are employed, are:

1. Convenience: selection of those who are easily available and willing to respond.
   - Snowball sampling is often used for hard-to-reach cases (for example those with a stigmatising health condition) and involves locating an initial respondent of interest and asking them to identify others, who can then each nominate further potential candidates (Magnani et al. 2005).
2. Chance: attempt to avoid accusations of bias by introducing a chance factor.
   - Quota sampling (Scott et al. 2013) involves selecting a pre-determined number of respondents in various categories (e.g. by age group and gender), usually to match population proportions. Selection typically takes place at a convenient location (e.g. interviewing individuals walking to market or waiting at a clinic).
   - Transect sampling (Leslie et al. 2009) is a process in which the researcher takes a central position in a community, selects a random direction (traditionally by spinning a bottle) and then selects respondents who live along that direction.
3. Purposive: sample determined by researcher using knowledge of the population and context to meet stated objectives (Palinkas et al, 2013)
   - Typical (modal instance) sampling aims to select a sample that is reasonably representative of a given type of respondent.
   - Heterogeneity (diversity/maximum variance) sampling aims to include the full range of potential responses.

- Stratified purposive sampling is similar to quota sampling, but with the specific respondents in each category determined by the researcher.

Probability sampling

One special group of sampling methods involves the use of '*probability sampling*'. This is the approach adopted in large scale sample surveys such as the Demographic and Health Surveys, World Health Surveys and the Multiple Indicator Cluster Surveys, and it is the reason that they are seen as providing *unbiased* estimates of key population parameters such as infant mortality rates. The following additional definitions apply:

- Probability Sample: A sample selected in such a way that the probability of selection could in theory be calculated for every member of the sample.
- Sampling frame: List of entities (or groups of entities) used to select the sample.
- Population Parameter: Quantitative information about the sampling population.
- Sample statistic: Quantitative information about the sample.
- Estimation: The use of a sample statistic to estimate a population parameter.

The first definition imposes a strict criterion for distinguishing between *probability samples* and *non-probability samples*. Because the former are widely regarded as least likely to be influenced by sampling errors, there is a tendency for researchers to proceed as if they analysing probability samples when they are not. For example, those conducting market research surveys or political opinion polls will typically adopt methods of analysis which are appropriate to probability samples even though they have not followed sampling procedures that meet the above criterion.

The nature of the criterion is perhaps most easily understand if we examine the distinction between a *chance* sample and a *random* sample. These words are often used interchangeably by those who have no knowledge of sampling theory but in fact have very different meanings. Consider a large meeting taking place in a conference hall. Researchers selecting a sample of ten individuals might stand by the door and interview the first ten people to emerge. If they were interested in gender differences, they might interview the first five women and the first five men. If they were concerned that those leaving first were likely to have different views than those less anxious to leave, they might sample every tenth person until they had reached their target. No matter what additional strategies the researchers introduce, they cannot claim that they have obtained a probability sample. There is no way that they can calculate the probability of selection for any given member of their sample. On the other hand this would be possible if, for example, the names of those attending were written on pieces of paper and placed in a bag, with ten names being selected from that bag. We would then say that every person attending the meeting had an equal probability of being selected, with that probability being equal to ten divided by the total number attending. This would be an example of random as opposed to chance sampling.

The key additional element is the use of a sampling frame. In the above, we can think of the names as being on a list which is then cut up and put into the bag, which simply serves as a mechanisms for selecting names *at random* from the list. More traditionally, the sample would be selected from the list using a computer generated *table of random numbers* or, more recently, using a mobile phone app. In some cases such lists will already exist. For example, there may be a list of all licenced doctors or pharmacists practicing in a city or a list of all rural public health facilities in a given region. In the latter case, if we wished to obtain a sample of all the doctors working in such facilities we could take a random sample of facilities, ask each of the sampled facilities to prepare a list of their doctors and then randomly sample from those lists. This would be an example of *two-stage probability sampling*. Rural household surveys often follow a similar procedure, sampling villages at the first stage and households within the sampled villages at the second.

The most common application of probability sample surveys is *estimation* of *population parameters* using *sample statistics*. Most often this involves the estimation of population means (for example mean number of antenatal visits) or proportions (proportion experiencing a fever in previous two weeks), using the equivalent measures calculated using data from the sample. As indicated above, if it can be determined that the appropriate procedures have been followed in conducting the survey, such estimates are widely regarded as unbiased. A further advantage of probability sampling, which applies to no other approach, is that it is also possible to estimate mathematically the magnitude of the sampling error, the risk that the sample is unrepresentative of the population, from the survey data. This allows statements as to the confidence with which the value of the estimated parameter can be assumed to lie between lower and upper bounds. For example, it might be asserted the proportion experiencing a fever in the previous two weeks (P) may be assumed, with 95% confidence ([NCCMT](#)), to lie within the bounds $P - k_1 \times se$ and $P + k_2 \times se$. Where k_1 and k_2 are known constants and se is the sampling error calculated from the survey data. Such bounded estimates are not only of value to researchers, for example when they are trying to assess changes over time due to an intervention, but also to decision makers attempting to assess the potential costs and benefits of such an intervention when considering a possible change in policy. Perhaps surprisingly, relatively small probability samples can provide reasonably precise bounded estimates for very large populations - in fact the size of the sampling population does not significantly affect the precision of the estimate as long as it is much greater than the size of the sample.

It should however be noted that probability sampling need not be restricted to studies which seek estimates of population parameters. When sample sizes are very small (perhaps less than ten observations), it may well seem preferable to use some form of purposive sampling, given that the risk of selecting an obviously inappropriate probability sample may be considerable. However, adopting a probability sampling approach can be very useful in larger scale qualitative studies that wish to argue that their findings can be generalised to a wider population. In this instance the primary advantage is that samples selected using this approach will be less open to challenges on the grounds of bias. As long as researchers can claim, and hopefully demonstrate, that they have followed the standard procedures, even critics should be willing to accept that there has been no attempt to subvert the study findings by the deliberate selection of a biased sample.

Against the above advantages must be set the potential costs associated with the need to obtain or construct a sampling frame. This may even raise difficult conceptual issues in terms of the entities to be included or excluded. For example, how should we define a private healthcare facility? Should we list only accredited facilities, even if the great majority of the population use unqualified providers who work from their own homes? Should we include traditional or religious healers? There will often also be considerable practical difficulties. How can we possibly identify all the unaccredited providers in a given area? How can we possibly construct a sampling frame that will allow us to study the healthcare needs of transient migrant workers who seek employment as day labourers? Existing lists, even of health facilities and communities are often outdated and incomplete. Construction of a new list may be difficult, time consuming and expensive, though researchers have found many ingenious ways to address this issue, often by adopting area sampling approaches ([Myatt et al. 2005](#); [Bennett et al. 1991](#)), a technique which has become much more readily available with the widespread use of mobile phones that can link to the Geographical Positioning System (GPS).

References

Barnett, I., Yosellina, Sulistyo, S., Befani, B., KariSari, K., Sharmin, S. and Dewi, D. (2016). *Mixed-Method Impact Evaluation of a Mobile Phone Application for Nutrition Monitoring in Indonesia*. IDS Evidence Report 200. Institute of Development Studies UK. http://www.ids.ac.uk/publication/mixed-method-impact-evaluation-of-a-mobile-phone-application-for-nutrition-monitoring-in-indonesia

Barahona, C., & Sarah Levy, S. (2002). *How to generate statistics and influence policy using participatory methods in research.* Statistical Services Centre Working Paper: University of Reading. www.reading.ac.uk/ssc/n/TIP%20Archive%20CD/Methodology/How_to_generate_stats_and_influence_policy.pdf

Bennett, Steve, Tony Woods, Winitha M. Liyanage and Duane L. Smith (1991). A Simplified General Method for Cluster-Sample Surveys of Health in Developing Countries. *World Health Statistics Quarterly* 44:98-106.

http://apps.who.int/iris/bitstream/10665/47585/1/WHSQ_1991_44%283%29_98-106_eng.pdf?ua=1

Bhuiya, A., Hanifi, S. M. A., Urni, F., & Mahmood, S. S. (2009). Three methods to monitor utilization of healthcare services by the poor. *International Journal for Equity in Health* 8.

Bishai, David, Ligia Paina, Qingfeng Li, David H Peters and Adnan A Hyder (2014). Advancing the application of systems thinking in health: why cure crowds out prevention. *Health Research Policy and Systems* 12:28. www.biomedcentral.com/content/pdf/1478-4505-12-28.pdf

Boslaugh, Sarah (2007). *Secondary Data Sources for Public Health: A Practical Guide*. Cambridge University Press.

www.academia.edu/1630213/An_introduction_to_secondary_data_analysis

Bowden, A, J A Fox-Rushby, L Nyandieka, and J Wanjau (2002). Methods for pre-testing and piloting survey questions: illustrations from the KENQOL survey of health-related quality of life. *Health Policy and Planning* 17(3):322-330.

Bowen, Glenn A (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal* 9(2):27-40.

www.academia.edu/7849680/Document_Analysis_as_a_Qualitative_Research_Method

Carvalho, S. and White, H. (1997). *Combining the quantitative and qualitative approaches to poverty measurement and analysis*. World Bank Technical Paper 366. The World Bank, Washington D.C.

Chambers, R. (2007). *Who Counts? The Quiet Revolution of Participation and Numbers*. Working Paper 296: Institute of Development Studies. https://www.ids.ac.uk/files/Wp296.pdf

Chambers, Robert (2006). *Poverty Unperceived: Traps, Biases and Agenda*. Working Paper 270. Institute of Development Studies. UK. www.ids.ac.uk/files/Wp270.pdf

Deitchler, Megan, Joseph J Valadez, Kari Egge, Soledad Fernandez and Mary Hennigan. 2007. A field test of three LQAS designs to assess the prevalence of acute malnutrition. *International Journal of Epidemiology* 36:858-864.

ESRC (undated), *ReStore: A Sustainable Web Resources Repository*. Economic and Social Research Council UK. http://www.restore.ac.uk/lboro/research/software/caqdas.php

Finlay, Linda (2002). Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative Research* 2:209. www.uk.sagepub.com/braunandclarke/study/SAGE Journal Articles/Ch 2. Finlay.pdf

Galvin, K. T., Petford, N., Ajose, F., & Davies, D. (2011). An exploratory qualitative study on perceptions about mosquito bed nets in the Niger Delta: what are the barriers to sustained use? *Journal of Multidisciplinary Healthcare* 4:73-84. https://www.dovepress.com/an-exploratory-qualitative-study-on-perceptions-about-mosquito-bed-net-peer-reviewed-article-JMDH

Greenhalgh. Trisha and Rod Taylor (1997). Papers that go beyond numbers (qualitative research). *BMJ* 315(20):740-743.

Hemingway, Pippa and Nic Brereton (2009). *What is a systematic review?* Hayward Group Ltd. (http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/syst-review.pdf

Holland, J. (Ed.) (2007). *Tools for Institutional, Political, and Social Analysis of Policy Reform A Sourcebook for Development Practitioners*. Washington: World Bank. http://siteresources.worldbank.org/EXTTOPPSISOU/Resources/1424002-1185304794278/TIPs_Sourcebook_English.pdf

Hoshaw-Woodard, Stacy. 2001. *Description and comparison of the methods of cluster sampling and lot quality assurance sampling to assess immunization coverage*. WHO. www.who.int/vaccines-documents/DocsPDF01/www592.pdf

Hsieh, Hsiu-Fang and Sarah E. Shannon (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15:1277-1288. http://www.iisgcp.org/pdf/glssn/Supplemental_Reading_on_Coding_2.pdf

Kanbur, Ravi and Paul Shaffer (2005). *Epistemology, Normative Theory and Poverty Analysis: Implications for Q-Squared in Practice*. Q-Squared Working Paper No. 2. University of Toronto. http://isiarticles.com/bundles/Article/pre/pdf/21648.pdf

Kanbur, Ravi (2003). Q-squared? A commentary on qualitative and quantitative poverty appraisal. In *Q-squared: qualitative and quantitative poverty appraisal,* ed. Ravi Kanbur, pp. 1-21. Permanent Black, Delhi. http://www.arts.cornell.edu/poverty/kanbur/qqz.pdf

Kuznetsov, Vladimir N, Andrej M Grjibovski, Andrej O Mariandyshev, Eva Johansson, Donald A Enarson and Gunnar A Bjune. Hopelessness as a basis for tuberculosis diagnostic delay in the Arkhangelsk region: a grounded theory study. *BMC Public Health* 13:712. www.biomedcentral.com/content/pdf/1471-2458-13-712.pdf

Lagarde, M. (2012). How to do (or not to do) . . . Assessing the impact of a policy change with routine longitudinal data. *Health Policy and Planning* 27:76-83. http://heapol.oxfordjournals.org/content/early/2011/01/27/heapol.czr004.full.pdf

Lamba S, Nityananda V (2014). Self-Deceived Individuals Are Better at Deceiving Others. *PLoS ONE* 9(8).

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104562#pone-0104562-g002

Leslie, Toby, Dr Nader Mohammed, Hafiz Omar, Haroon Ur Rasheed, Eveline Klinkenberg, Frits van der Vorst (2009). Assessment of the Present Role of the Private Sector in Malaria Diagnosis, Treatment and Control in Afghanistan. *Afghanistan Annual Malaria Journal* 1:89-113. http://moph.gov.af/Content/Media/Documents/AAMJ2812201015030182.pdf

Lover, A. A., Sutton, B. A., Asy, A. J., & Wilder-Smith, A. (2011). An exploratory study of treated-bed nets in Timor-Leste: patterns of intended and alternative usage. *Malaria Journal* 10.

Lucas, Henry, Shijun Ding and Gerald Bloom (2009). What do we mean by 'major illness'? The need for new approaches to research on the impact of ill-health on poverty. In *Health and social protection: experiences from Cambodia, China and Lao PDR*. Bruno Meessen, Xiaomei Pei, Bart Criel and Gerald Bloom, eds. ITGPress, Antwerp. http://www.eldis.org/vfile/upload/1/document/0809/What do we mean by major illness - The need for new approaches to research on the impact of ill-health on poverty.pdf

Magnani, Robert, Keith Sabin, Tobi Saidela and Douglas Heckathorn (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS* 19 (suppl 2):S67-S72. http://journals.lww.com/aidsonline/Abstract/2005/05002/Review_of_sampling_hard_to_reach_and_hidden.9.aspx

Marsland, N., Wilson, I., Abeyasekera, S., and Kleih, U. (1998). *A Methodological*

Framework for Combining Quantitative and Qualitative Survey Methods - Background

Paper: Types of Combinations. Report written for DFID Research project R7033. http://www.reading.ac.uk/ssc/resources/Docs/QQA/qqa.pdf

Martineau, Tim, Tolib Mirzoev, Stephen Pearson, Bui Thi Thu Ha, Qian Xu, K V Ramani and Xiaoyun Liu (2015). Coherence between health policy and human resource strategy: lessons from maternal health in Vietnam, India and China. *Health Policy and Planning* 30:111-120.

Mruck, Katja; Breuer, Franz (2003). Subjectivity and Reflexivity in Qualitative Research—The FQS Issues. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 4(2).

www.qualitative-research.net/index.php/fqs/article/view/696/1504

Murray, C. J. L. (2007). Towards good practice for health statistics: lessons from the Millennium Development Goal health indicators. *Lancet* 369:862-873.

Myatt, Mark, Teshome Feleke, Kate Sadler, and Steve Collins. 2005. Field trial of a survey method for estimating the coverage of selective feeding programmes. *Bulletin of the World Health Organisation* 83(1):20-26. www.who.int/bulletin/volumes/83/1/20.pdf

Neupane, Sunisha, Willem Odendaal, Irwin Friedman, Waasila Jassat, Helen Schneider and Tanya Doherty (2014). Comparing a paper based monitoring and evaluation system to a mHealth system to support the national community health worker programme, South Africa: an evaluation. *BMC Medical Informatics and Decision Making* 14:69. http://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-14-69

Nyandieka, Lilian N, Annabel Bowden, John Wanjau, and Julia A Fox-Rushby (2002). Managing a household survey: a practical example from the KENQOL survey. *Health Policy and Planning* 17(2):207-212

Nzinga, Jacintha, Patrick Mbindyo, Lairumbi Mbaabu, Ann Warira and Mike English. (2009). Documenting the experiences of health workers expected to implement guidelines during an intervention study in Kenyan hospitals. *Implementation Science* 4:44. http://www.implementationscience.com/content/4/1/44

Palinkas, Lawrence A, Sarah M. Horwitz, Carla A. Green, Jennifer P. Wisdom, Naihua Duan and Kimberly Hoagwood (2013). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health* 42(5). https://www.researchgate.net/publication/258315317_Purposeful_Sampling_for_Qualitative_Data_Collection_and_Analysis_in_Mixed_Method_Implementation_Research

Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011). Mixed Method Designs in Implementation Research. Administration and Policy in Mental Health and Mental Health Services Research 38:44-53. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3025112/

Peters, D. H., Ahmed, A., Noor, A., Singh, L. P., Kakar, F. K., Hansena, P. M., et al. (2007). A balanced scorecard for health services in Afghanistan. *Bulletin of the World Health Organization* 85(2):146-151. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2636277/

Qualquant (2016). *Qualquant: Use the Right Tool*. National Science Foundation. http://qualquant.org/

Rifkin, Susan B (1996). Rapid Rural Appraisal: Its use and value for health planners and managers. *Public Administration* 74:509-526. http://www.academia.edu/3664632/RAPID_RURAL_APPRAISAL_ITS_USE_AND_VALUE_FOR_HEALTH_PLANNERS_AND_MANAGERS

Rog, Morgan, Bonnielin Swenor, Luis C Cajas-Monson, Wilson Mchiwe, Steven Kiboko, Harran Mkocha and Sheila West (2011). A cross-sectional survey of water and clean faces in trachoma endemic communities in Tanzania. *BMC Public Health* 11:495. www.biomedcentral.com/content/pdf/1471-2458-11-495.pdf

Rowe, A. K. (2009). Potential of Integrated Continuous Surveys and Quality Management to Support Monitoring, Evaluation, and the Scale-Up of Health Interventions in Developing Countries. *American Journal of Tropical Medicine and Hygeine*, 80(6):971-979. http://www.ajtmh.org/content/80/6/971.long

Scott, Jennifer, Sarah Averbach, Anna Merport Modest, Michele Hacker, Sarah Cornish, Danielle Spencer, Maureen Murphy and Parveen Parmar (2013). An assessment of attitudes toward gender inequitable sexual and reproductive health norms in South Sudan: a community-based participatory research approach. *Conflict and Health* 7(24). www.biomedcentral.com/content/pdf/1752-1505-7-24.pdf

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education* 5(4), 541-542. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3886444/pdf/i1949-8357-5-4-541.pdf

Uganda (2012). *mTrac: Information for Better Health*. Ugandan Ministry of Health. http://www.mtrac.ug/video-tags/uganda

Wallerstein, N. B., & Duran, B. (2006). Using Community-Based Participatory Research to Address Health Disparities. *Health Promotion Practice* 7:312-323.

Walshe C, Ewing G, Griffiths J. (2012). Using observation as a data collection method to help understand patient and professional roles and actions in palliative care settings. *Palliative Medicine* 26(8):1048-1054.
https://www.researchgate.net/publication/51898162_Using_observation_as_a_data_collection_method_to_help_understand_patient_and_professional_roles_and_actions_in_palliative_care_settings

World Bank (2003). *Measuring Living Standards: Household Consumption and Wealth Indices.* Quantitative Techniques for Health Equity Analysis—Technical Note #4. World Bank. http://siteresources.worldbank.org/INTPAH/Resources/Publications/Quantitative-Techniques/health_eq_tn04.pdf

Chapter 8: Quantitative methods

Henry Lucas
Institute of Development Studies

1.   Introduction

As discussed in Chapter 7, in order to ensure a degree of independence, implementation research will typically not be funded from the intervention budget and the level of funding will almost always be relatively limited. Implementation researchers are therefore not in a position to insist that the data they need to meet their specific objectives should be made available within the general monitoring and evaluation (M&E) system. However, where possible, they should aim to be involved in the design of that system and may be able to negotiate modifications that serve their purposes. This may be possible either because those modifications are also seen as valuable to those with overall responsibility for the implementation - for example supporting operations research activities - or because additional resources from the implementation research budget are made available to complement those allocated to M&E within the intervention.

The above implies that implementation researchers will rarely be able to embark on independently managed and funded large-scale primary data collection activities but will have to rely mainly on the intervention M&E system, special studies using the 'qualitative methods' described in Chapter 8 and secondary sources. The key responsibility in this case is to adopt a systematic approach to determine the quality of the data to be analysed. This will be an ongoing challenge, given the tendency for data quality to vary over time, possibly improving initially as innovative systems for data collection are introduced and the enthusiasm of those involved is stimulated by access to new equipment and training workshops, but often deteriorating as that enthusiasm declines and systems fail to work as intended.

An obvious starting point when assessing data quality is the existence and completeness of those data. Missing information on facilities, providers, patients, etc. not only limits the scope of the analysis that can be undertaken, it typically biases the findings of that analysis. As a rule it will be the less well-resourced, less well-managed, most remote locations that are most at risk of providing incomplete data. Failure to recognise this trend can lead to a seriously over-optimistic view of intervention progress. Given that data are available, the statistical agency of the European Commission (Eurostat 2007) defines data quality as having five additional desirable attributes:

1. Accuracy;
2. Timeliness;
3. Comparability;
4. Coherence;
5. Accessibility and clarity.

Accuracy - essentially whether data reflect the true value of a given quantity - is obviously very difficult to test. However, we can check for obvious outliers, values that are almost certainly too large or too small. This can be a very important check, as many statistical procedures are highly sensitive to outliers, which can seriously distort the findings of any analysis. If we have time series data, excessive changes between one period and the next may also indicate measurement or recording problems. In some cases, issues may become apparent if we calculate rate or ratio indicators. For example, is the number of patients seen per day per doctor plausible? Are the recorded financial data compatible with patient numbers? In addition, by examining the frequency distributions of selected data items, as discussed below, it may

be possible to determine if initial assumptions about those items have proved valid. For example, attempts to assess patient satisfaction using a scale often result in a distribution in which almost no patients chose the lowest points on that scale. This should raise questions as to whether the scale we are using is an accurate reflection of patient attitudes.

Timeliness reflects the delay between the occurrence of an event or phenomenon and the availability of the associated data items. It is relevant in terms of both routine data systems and the intervention M&E system, which should be providing time series data that allow us to track implementation progress and link intervention inputs to outputs and outcomes. For example, a training workshop at a given point in time may be intended to result in improved staff performance, which is expected to produce better health outcomes by some future date. Assessment of the extent to which this sequence of events has taken place may be complicated by excessive delays in the availability of data from some sources, given that, as with missing data, such delays will often be associated with facilities or agencies that are performing less well.

Comparability relates to differences in concepts and measurement tools/procedures between sources - e.g. facilities, geographical locations, etc. - or over time. This can be a particularly serious problem in research on health systems, where different providers often choose to specify their own diagnostic and treatment protocols. Some unqualified providers may record all patients with a fever as suffering from malaria, while others rely on a range of rapid diagnostic tests. Some hospitals may record a diagnosis of tuberculosis based only on a chest X-ray while others require identification of Mycobacterium tuberculosis from clinical specimens. Additional issues arise if definitions are modified over time, perhaps as a result of the intervention itself. For example, faced with restrictions on outpatient costs by health insurance schemes, providers may simply vary their accounting procedures such that costs are transferred to inpatient departments. Such possibilities need to be carefully considered to avoid misinterpretation of apparent trends over time.

Where analysis involves the combination of data items from different sources it will be necessary to assess the extent to which there is *coherence* between those items. For example, to determine the extent to which some conditions remain untreated in the population it may be necessary to combine aggregates calculated from facility routine data systems with estimates of the prevalence of those conditions based on data from existing surveys. Those responsible for compiling these two sources will typically have used very different concepts, methodologies and instruments because they had distinct objectives. To the extent possible, any analysis must evaluate and try to address the implications of these differences.

Accessibility and clarity are perhaps most usefully understood as denoting the extent to which a researcher has the required level of understanding about the true nature of the data they intend to analyse. At a minimum this implies a careful review of the necessary 'metadata' - documentation that describes how the data are intended to be collected, compiled and stored - assuming these are available. However, it will often be clear that this documentation has limited relevance in terms of how those responsible for these activities proceed in practice. For example, heath staff tasked with introducing new procedures for collecting and recording patient data will typically soon find ways to reduce the time and effort required for an activity that they will often regard as a pointless addition to their workload. They may make assumptions as to the personal characteristics of the patient rather than ask the appropriate questions, or decide to enter data into the computer at the end of the day rather than 'waste' time after each patient visit as intended. The potential for misinterpretation if such issues are not addressed is evident. It may be frustrating to discover that an intended analysis cannot be undertaken as planned because the required data are not as originally assumed, but, as emphasised in Chapter 7, implementation research demands the highest standards of integrity and that includes ensuring that data limitations are thoroughly examined and addressed.

2.   Rapid surveys

One additional methodological tool that can prove very useful in implementation research is the rapid survey. This description is usually applied to relatively small-scale surveys, typically of around 200 subjects or less, which aim to collect a very limited number of items of quantitative data over a short time period, say 5–10 days, that can be analysed and interpreted within at most a few weeks. Rapid surveys can target a variety of populations including facility records, health staff, patients, households and individuals. They adopt the probability sampling approach described in Chapter 7, and can therefore be analysed using statistical inference procedures that provide unbiased estimates of population 'parameters' (quantitative information) and reliable estimates of error bounds on those parameters. They should not be used to address more complex questions, for example the detailed operation of new incentive schemes or implications of new mechanisms for reimbursement of user fees. It is often assumed that surveys can be used to 'find out about' a policy question. In fact, the successful planning and implementation of a rapid survey typically requires that a great deal is already known, both about the population to be surveyed, and about the subject matter under investigation.

Rapid surveys are usually cross-sectional but may also be used to track changes over time (Tipping and Segall 1996). The small sample size and limited number of data items greatly reduce the administrative and logistic burdens associated with large-scale, multi-topic sample surveys, particularly those associated with the recruitment and training of field staff. This does not imply that such surveys should be undertaken without due consideration of the implications in terms of resource allocation. Though they are sometimes described as 'lightweight', it is essential that they are designed and implemented with all the rigour that should be applied to any study that intends to claim the respect that is reserved by many policymakers for findings derived from traditional statistical surveys. The range of tasks to be undertaken is identical to that required for a large-scale survey, even if the content of each is much more limited:

- Questionnaire design;
- Sample design;
- Mapping/listing to create the sampling frame;
- Preparation of fieldwork manuals;
- Recruitment of field staff;
- Training of field staff;
- Field enumeration and supervision;
- Transport and communications;
- Data preparation and processing;
- Computer analysis.

A number of the above require human resource management skills that some researchers either lack or are reluctant to practise. Again as with large-scale surveys, a key point to bear in mind is that not all of those involved in a survey can be assumed to have a direct personal stake in achieving a successful outcome. Without effective management and supervision, supported by a system of incentives and penalties, many will not perform to the standard required. Apart from such practical issues, it is also necessary to give some thought to the legal and administrative context within which surveys are undertaken. Are there laws that prohibit data being taken from a patient record or doctors providing information on the health status of an individual? If these activities are legal, is it necessary to obtain permission from a relevant administrative agency before undertaking them? Even if we have such permission, perhaps only because that agency wishes to encourage the intervention with which we are associated, we should still consider if we are abiding by the ethical criteria described in Chapter 6. As this is intended as a practical guide, we would also advise consideration of any

potential political implications of our survey. Are we addressing a sensitive issue? Might some stakeholders be concerned that we are gathering information that might be used to their disadvantage? What are the possible implications in terms of our overall research activities?

Another set of potential constraints that may impact on the quality of the survey are those relating to the targeted respondents. Can we assume, given that we have the appropriate permissions, that they will be cooperative? If you have undertaken the detailed stakeholder analysis discussed in Chapter 5, the findings from that analysis should provide information that will help you make such a judgement. Does what you know of your intended respondents indicate that they may have reasons - guilt, embarrassment, suspicions about your motives - to be concerned about providing you with data? Might they be irritated by what they see as an interruption to their normal activities? For example, in many countries busy frontline health workers often tend to regard all record-keeping as a largely pointless chore that takes time away from patient care. On the other hand, might their desire to be helpful - perhaps because they regard you as a high-status individual, or simply out of a natural tendency to be polite - lead them to provide data that they know to be unreliable and/or incomplete, rather than risk disappointing you?

This raises another issue. Even if respondents are cooperative, do they have access to the data you require? Do those data relate to current knowledge that they almost certainly possess or memories that may have become less reliable over time? Will they need to consult records? If so, do such records exist and are they complete and reliable? Finally, it is important to remember that one major disadvantage of questionnaire surveys is that it is very difficult in practice to ensure that every question will be interpreted in precisely the same way by *all* respondents. As a first priority, we should, if possible, try to ensure that the enumerator and respondent share fluency in a common language and that the questionnaire has been translated into that language (standard practice is to translate the questionnaire and then translate back for comparison with the original). Obviously, every effort should be made to avoid ambiguity and complexity in language. One useful approach is to deliberately try to identify any remotely possible way in which questions might be open to misinterpretation. Health-related surveys raise particular issues, in that researchers sometimes casually use technical terms that seem commonplace to them but that may be interpreted quite differently by some sections of the surveyed population. For example, the term 'inpatient care' is usually taken to imply at least one night spent in hospital, but could be seen as applying to any individual who has received treatment in a hospital inpatient department, for example reclining on a bed to receive a saline drip. Similarly, to a researcher the word 'doctor' may signify a qualified, licensed professional. In a remote village the same word may be used for an unqualified traditional healer.

Sampling designs
The sampling designs used in rapid surveys, as in the great majority of large-scale surveys, are based on the combination of a relatively limited number of elements:

- Simple random sampling;
- Systematic (list) sampling;
- Stratified sampling;
- Sampling with probability proportional to size;
- Cluster sampling.

The differences between these procedures can best be understood by considering a simple example. Suppose we wish to estimate the proportion of hospital clinical staff who have understood the basic principles of an innovative procedure following a one-week training course. If we have a list of all the staff in the hospitals, we could take a random sample and calculate the proportion of our sample who can answer a few simple questions about the

procedure. That could be used as an unbiased estimate of the proportion of all staff that would have been able to answer those questions, which we could interpret as the proportion with a good knowledge of the procedure. There are two slightly different types of random sample. In simple random sampling (SRS), we would select members of staff sequentially from the full list, which allows for the possibility that we may select the same member more than once. In simple random sampling without replacement (SRSWOR), we again sample sequentially, but this time excluding any member previously selected.

SRS is used as a standard sample design to which others are compared. It allows calculation of simple estimates of the required sample size for a given level of *precision* (the size of the lower and upper error bounds for the estimate). Thus:

1. A sample of size 100, selected using SRS allows estimation of a proportion to a precision of +/-10 per cent with 95 per cent confidence;
2. A sample of size 400, selected using SRS allows estimation of a proportion to a precision of +/-5 per cent with 95 per cent confidence (note that improving precision by a factor of two requires increasing the sample size by a factor of four).

We can interpret '95 per cent confidence' as implying that only 5 per cent (1 in 20) of such samples would be so misleading as to result in an estimated proportion that was further away from the true population parameter. We simply assume that we have not been so unlucky as to have chosen one of those samples[iv].

In theory, SRS can be used as the reference to calculate the *efficiency* of any given sample design:

Efficiency = precision of SRS/precision using alternative design and same sample size

However, typically we do not have sufficient information to estimate efficiency but may simply be aware that one design is almost certainly more efficient than another. This is important because even though we may not be able to calculate the cost of achieving a given level of precision, a more efficient design can deliver increased precision for the same cost - i.e. it will probably result in a better estimate.

As indicated in Chapter 7, taking a random sample requires repeated use of a set of random number tables, a computer program or more recently a mobile phone app. Systematic sampling is a simpler approach that involves selecting a starting point on the list at random and then sampling every kth entry, where k is equal to the total number of entries divided by the required sample size (k=N/n). When the end of the list is reached, the process continues from the first entry. Except in rare cases where the list happens to follow a pattern that increases the risk of a biased sample, it can be shown that this procedure produces unbiased estimates with sampling errors that are at least as small as those from a random sample. It can in theory be more efficient than SRS if the list is ordered by a variable that is related to staff performance, for example if the names are listed by hospital, because there is less risk of selecting an unrepresentative sample - that is, one that does not include staff from all hospitals. However, as indicated above, it will typically be impossible to estimate the extent of the gain in efficiency.

If we suspect that knowledge of the procedure among staff may differ substantially between hospitals, we can increase the precision of estimation - i.e. narrow the gap between the lower and upper bounds on our estimate - by ensuring that our sample must include staff from every hospital. For example, in each hospital we might take a constant proportion (k=n/N) of staff members. This would be a stratified sample, with each hospital being a separate stratum. This sampling design results in a reduced sampling error compared to random sampling because we have excluded the risk of selecting samples that were mainly from under- or over-

performing hospitals, samples which would have under- or over-estimated the proportion of knowledgeable staff in the total population. The unbiased estimator of this parameter is calculated as a weighted sum of the proportions in each hospital (p_i), where the weights are equal to the number of staff in each hospital (N_i) divided by the total number of staff (N), i.e.

$$P = \Sigma\ N_i/N \times p_i.$$

Stratification by a range of other variables, for example gender, age or grade of staff, etc., might similarly be used to reduce the sampling error, if it were suspected that they were also associated with differences in staff knowledge. A basic principle is that the more information we have about our target population, the easier it will be to develop an efficient sampling design. It is important to understand that a stratified sampling design is intended to provide better estimates of overall population parameters. In the above example, we would be calculating statistics for individual hospitals and it will be tempting to compare, for example, average performance levels between those hospitals. However, that was not the purpose of the sample design and we will usually find that we simply do not have sufficient observations in each hospital to make such comparisons reliably.

For the estimation of a range of key population parameters, including totals, averages and rates, large entities - villages or urban districts with large populations, hospitals with a large number of inpatient beds, diseases with a high prevalence rate - are obviously very important in terms of their contribution to those parameters. In the above example, failing to obtain data from a small district hospital would have little impact on our overall estimate of the proportion of staff with adequate knowledge. However, failing to include staff from the largest national hospital could easily make a substantial difference to our estimate. One way to address this issue would be to stratify staff by size of hospital. As an alternative, the probability P of including a staff member of a given hospital in our sample could be made proportional to the total number of staff in the hospital, for example for all staff members in hospital i, the probability of being selected is:

P(i) = number of staff in hospital(i) / number of staff in all hospitals

This approach, called sampling with 'probability proportional to size' (PPS), increases efficiency by increasing the probability of inclusion in the sample for staff from large hospitals, thus decreasing the risk of taking a sample that excludes staff from these hospitals.

The PPS design is most commonly used in *cluster sampling* (Bennett et al. 1991), which is also referred to as two-stage sampling. In our example the hospitals can be regarded as 'clusters' of staff. If we decided that it would be too expensive, for example in terms of travel and accommodation costs, to send enumerators to every hospital within our study area, we might decide to (1) select a sample of hospitals and then (2) select a sample of staff within each of those hospitals. A common sampling design would be to use PPS to sample hospitals and systematic sampling to sample staff within each hospital. Cluster sampling almost always involves a loss of efficiency for a given overall sample size. As not all hospitals will be surveyed, if there are differences between them this design introduces a risk of selecting a sample of hospitals that is not representative. This risk increases as the number of hospitals in the sample decreases. Cluster samples typically need to be two to ten times as large as an SRS to achieve the same precision.

Estimation of sampling errors

In each of the above designs, the sample selected is simply one of the many that might have been selected using the same design and with the same sample size. The sampling error of an estimate is essentially a measure of the variability between all possible values of that estimate that might have been obtained from different samples. One way to reduce that variability is to increase the sample size but that will imply a higher cost. The other is to improve

the sample design, adopting sampling procedures that attempt to maximise, for a given sample size, the proportion of possible samples that will provide estimates that are close to the population parameter. We can never ensure that the sample that we do obtain meets this requirement, but using probability sampling we can make a reasonable estimate of the sampling error, which determines the risk of a 'bad' sample, and use this to modify the sample design or increase the sample size to ensure that it is less than some designated level - typically 1 in 20 or 1 in 100.

For a simple random sample (SRS) the sampling error can be estimated using:

$$se_{srs} = \sqrt{[\Sigma(x_i - \overline{x})^2/n]}$$

where x_i (i=1..n) are the sample values, $\overline{x}$ is the arithmetic average or mean of those values and n is the sample size. If we are willing to take a risk of 1 in 20, a remarkable mathematical result called the *Central Limit Theorem*, which can be derived from the basic definitions of probability theory, allows us to construct a 95 per cent confidence interval for the mean of the sampling population:

$$\text{Population mean} = \overline{x} \pm 1.96\ se_{srs}$$

Or a 99 per cent confidence interval:

$$\text{Population mean} = \overline{x} \pm 2.58\ se_{srs}$$

Note that the more confident we wish to be, the wider must our interval be. A similar formula can also be applied to confidence limits for proportions, as in the example discussed above.

Example: A study was designed to evaluate the effect of integrating ITN (insecticide treated bednet) distribution on measles vaccination campaign coverage in Madagascar (Goodson et al. 2012). A national cross-sectional survey was undertaken to estimate measles vaccination coverage, nationally, and in districts with and without ITN integration. To evaluate the effect of ITN integration, propensity score matching was used to create comparable samples in ITN and non-ITN districts. Relative risks (RR) and 95 per cent confidence intervals (CI) were estimated via log-binomial models. Equity ratios, defined as the coverage ratio between the lowest and highest household wealth quintile (Q), were used to assess equity in measles vaccination coverage.

National measles vaccination coverage during the campaign was 66.9 per cent (95 per cent CI 63.0-70.7). Among the propensity score subset, vaccination campaign coverage was higher in ITN districts (70.8 per cent) than non-ITN districts (59.1 per cent) (RR = 1.3, 95 per cent CI 1.1-1.6). Among children in the poorest wealth quintile, vaccination coverage was higher in ITN than in non-ITN districts (Q1; RR = 2.4, 95 per cent CI 1.2-4.8) and equity for measles vaccination was greater in ITN districts (equity ratio = 1.0, 95 per cent CI 0.8-1.3) than in non-ITN districts (equity ratio = 0.4, 95 per cent CI 0.2-0.8).

It should be emphasised that the above formula is only appropriate where the sample is selected using simple random sampling. There is a tendency for researchers to ignore this requirement and use the formula whatever the sample design adopted. As indicated in earlier chapters, the argument in this volume is that implementation research findings are too important for such disregard of established analytical procedures to be considered acceptable. To illustrate the problem, consider that rapid surveys will almost always adopt some form of cluster sampling. This implies that the above has to be modified to include a 'design effect', which measures the ratio of the sampling error of the cluster sampling design to that which would have resulted if an SRS design had been used.

Population mean = $\bar{x} \pm 1.96$ $se_{cluster}$

$$= \bar{x} \pm 1.96 \times \text{design effect} \times se_{srs}$$

The design effect will vary depending on the extent to which the clusters differ from each other. If this is large compared to the variability between the individuals within each cluster, the risk of sampling clusters that are unrepresentative is large and the design effect is large. In the above example, if the staff in some hospitals had all been well trained in the new procedure while in others training had been minimal, taking a cluster sample of a small number of hospitals would run a substantial risk of over- or under-estimating the overall level of staff proficiency.

The implications of a large design effect on the appropriate confidence limit bounds can be substantial. Table 1 below compares standard errors using the SRS formula with the appropriate standard errors for a clustered design where the clusters were villages. Note that the design effect varies considerably, from 1.13 for primary completion rates (limited between village variation because pupils travel to school) to 4.08 for improved drinking water (high proportion of variation between villages because this relates to a village level facility).

Table 1: Comparison of SRS and cluster sampling errors

Item	m	SRS se	Design Effect	Cluster se	m-2se	m+2se
Availability of ITN*	0.05	0.01	1.54	0.01	0.03	0.07
Iodised salt consumption	0.82	0.01	1.77	0.02	0.78	0.87
Improved drinking water	0.75	0.01	4.07	0.06	0.64	0.87
Primary completion rate	0.86	0.03	1.13	0.03	0.80	0.93
Attends secondary school	0.81	0.01	1.48	0.02	0.77	0.85

*Insecticide treated bednets

One reason for the inappropriate use of the SRS formula by research was the difficulty of calculating the correct sampling error, which often requires a considerable familiarity with the methods of theoretical statistics. However, such calculations can now be undertaken using well-established software packages such as STATA and SPSS, which require only that the researcher provide a detailed description of the sample design adopted.

The WHO Expanded Programme on Immunisation (EPI) surveys
The origin of the 'rapid survey' concept is often dated to the WHO '30 by 7' cluster surveys that were introduced in 1978 to obtain rapid, inexpensive but reasonably reliable estimates of child immunisation coverage (Lemeshow and Robinson 1985). The target population is subdivided into a complete set of non-overlapping 'clusters', usually defined by geographic boundaries (typically villages or urban districts). A sample of 30 of these clusters is taken with probability proportion to size (PPS) and then a 'quasi-random' sample of seven households with children in the relevant age range is selected within each of these clusters. Following this procedure, coverage estimates can be obtained that can be confidently assumed to be within ±10 per cent of the true value. The basic immunisation coverage survey instrument (the seven children sampled per cluster are typically recorded on a one-page document) usually records simply the cluster location, the age and sex of the selected child and their immunisation status. A similar methodology has been applied in rapid nutrition surveys, which have often been applied in emergency situations (Prudhon and Spiegel 2007). In this case it is usually recommended that the second-stage sample size should be increased to 30 children (SMART 2005).

The approach has attracted some criticism. [Turner et al. (1996)](#) focus on the lack of formal probability sampling of households within clusters. For example, one popular technique involves selecting a random direction from a central location within a village or urban district (traditionally by spinning a bottle). The households from the central point to the edge of the community in the chosen direction are listed, one is selected at random and then that household and its nearest neighbours are visited until the required seven children have been enumerated. None of the commonly used methods meets the basic requirement of probability sampling, that every eligible member of the target population has a known, non-zero chance of being selected. Simulation exercises suggest that the risk of sampling bias is substantially higher than in a conventional cluster sample. The paper suggests that a relatively simple modification can retain the advantages of the '30 by 7' design while ensuring a true probability sample. This involves: the production of a simple sketch map of each selected cluster; dividing this into segments of roughly equal size; selecting one segment at random; and interviewing all eligible members of the target population(s). This approach also addresses the common situation where surveyors attempt to gather information on multiple indicators (e.g. vaccination and childhood illness incidence rates) from the same sample.

[Myatt et al. (2005)](#) argue that while the PPS approach used in the '30 by 7' surveys may result in improved estimates overall, the associated tendency to sample areas of high population density may lead to a judgement that reasonable coverage has been achieved even where more remote, low-density areas have been severely neglected. They argue that this is of special concern in the case of feeding programmes, where a priority objective may be to identify such areas before children become severely malnourished. They describe an alternative approach which was first trialled in 2002 in the Mchinji district of Malawi where a district-wide feeding programme had been implemented. A 10km by 10km grid was overlaid on a map of the district. All those squares (quadrats) with more than 50 per cent of their area within the district were sampled. Communities nearest the centre of each quadrat were then sampled, with the sample size determined as the number that could reasonably be surveyed in a single day, based on the size of each community and the distance between them. All children in a community were screened to identify those suffering from malnutrition using a standard anthropometric criterion. Coverage in each quadrat was calculated as the proportion of malnourished children included in the feeding programme and overall coverage estimated by treating the quadrats as a stratum in a stratified sample. The survey was reported as proving simple, inexpensive and rapid, providing results within just ten days.

3.   Quantitative analysis

As discussed in Chapter 1, implementation research has two broad aims:

1. Understanding implementation processes, focusing on mechanisms that support or constrain those processes;
2. Communicating that understanding to the multiple stakeholders who may contribute to the integration of findings into current and/or future implementations.

Those stakeholders may include:

- The implementation team;
- Providers and other actors in the health sector;
- National and local policymakers/officials;
- NGOs and CBOs;
- Donor and other international agencies;
- Beneficiary communities;
- The general population.

A key issue is that very few of these stakeholders will have specialist knowledge of quantitative or qualitative methods. It is therefore of central importance that analysis and, most importantly, presentation of findings must be carefully considered to avoid potential misinterpretations that could lead to inappropriate responses. Emphasis needs to be placed on simplicity and interpretability - stakeholders need to both understand the information provided and interpret it correctly (Walker et al. 2007). In terms of quantitative analysis, this implies an emphasis on simple summary statistics such as:

- Counts, means, medians, ranges, percentiles;
- Rates, ratios and (for some stakeholders) risks;
- Frequency distributions, proportions and percentages.

This does not imply that complex analytical techniques are never appropriate; only that final communication of the analytical findings should meet the above criteria.

Designing analysis by purpose
A second important preliminary consideration is to clearly assess the primary objectives of any analysis - what specific issues are you trying to address? Implementation research is by nature intended not to simply describe specific implementations but to improve the process of implementation. For example, we might focus on:

Effectiveness: Research that aims to modify implementation procedures in order to improve the flow of benefits that result from a given level of resources. This is typically the primary aim of implementation research. It should also assess 'how effective' and 'for whom'?

Efficiency: Analysis that attempts to assess the implications of possible modifications to the implementation process in terms of the value of benefit flows relative to resource costs. The aim will be to improve the benefit/cost ratio.

Equity: Analysis of distributional issues, i.e. how are benefits and resource costs distributed, typically relating to population subgroups?

Sustainability: Focus on identification of essential inputs, potential constraints on their availability and other possible barriers to medium- and long-term sustainability.

The aim in this section is not to teach statistical methods but to consider, given the objectives described above, the most appropriate choice of methods in the context of implementation research. Five main areas are addressed:

1. Frequency distribution and summary statistics
2. Relationships and confounding variables
3. Subgroup analysis
4. Statistical models
5. Generalising from samples to populations.

A note on levels of measurement in quantitative studies
Variables are usually classified by their 'level of measurement':

1. Rational, e.g. weight of child, number of vaccinations;
2. Interval, e.g. temperature, some disability measures;
3. Ordinal, e.g. facility levels, quality of life indices;
4. Nominal, e.g. district names.

The level of measurement should determine the appropriate type of analysis - for example, using an ordinal dependent variable in a regression contravenes one of the assumptions of such models. Researchers often ignore such restrictions. However, as previously indicated, because the findings are explicitly intended to influence important implementation processes and to be interpreted and used by a wide variety of stakeholders, it is probably reasonable to set a higher standard in implementation research.

Distributions and summary measures

Implementation research data can be seen as distributions of the values of study variables over selected study populations. For example, we may consider the distribution of white blood cell counts across patients, the numbers of children under five across households or outpatient attendances on a given day across primary facilities. Analysis can be seen as the use of techniques intended to summarise those distributions and estimate the extent to which they are related. For example, in a sample of newborn children we might summarise the distribution of birth weights by calculating the frequency of low, normal and high weight births, classifying as 'normal' those in some standard range. If we also calculated the frequency of different education levels for the mothers of those children, we could estimate the strength of a possible relationship between these two variables.

This use of frequency distributions, which show the number of values of a given variable that fall in each of several non-overlapping (mutually exclusive) groups, for this purpose (table 2) has a number of advantages. They are useful for all types of variable, easy to explain and interpret for audiences without specialist knowledge and can be presented graphically (figure 1) and/or in different formats to aid interpretation.

Table 2: Provider education frequency distribution

Level of education of private providers	Frequency
Illiterate	106
Basic literacy	74
Primary school certificate	57
Secondary school certificate	11
Higher-level qualification	2
Total	250

Figure 1: Provider education bar chart
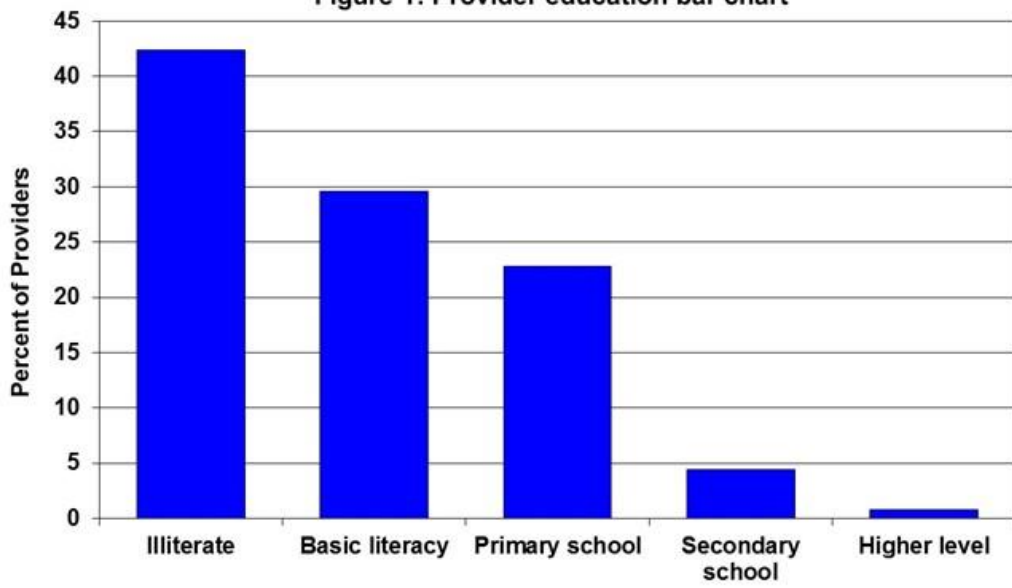
Frequency distributions provide an extremely useful approach to the presentation of large volumes of data. In the above example, information relating to 250 people has been used to construct one small table and, very importantly, no information has been lost in the process – that is, it would be possible to regenerate the original list of data values given the table. There are a number of interesting alternative ways of presenting the above data. We often, for example, calculate the 'relative frequency' (proportion, percentage) of data items that fall into a specific class. Again, to provide a slightly different perspective, we can 'cumulate' these percentages to show, for example, that 94.8 per cent of our population have at most a primary school certificate as in table 3.

Table 3: Alternative presentations of a frequency distribution

Level of education	Proportion	Percentage	Cumulative percentage
Illiterate	0.424	42.4	42.4
Basic literacy	0.296	29.6	72.0
Primary school certificate	0.228	22.8	94.8
Secondary school certificate	0.044	4.4	99.2
Higher-level qualification	0.008	0.8	100.0
Total	1.000	100.0	

Similarly we can experiment with different graphical displays. Figure 2 below shows the percentages as the segments of a pie chart. Note that the percentages are rounded to whole numbers. As a general rule, it makes sense to present data only to the degree of accuracy that (a) it warrants (estimates are almost always based on data that contain errors); and (b) makes the point we wish to make. Excessive precision (for example expressing numbers to more than one or two decimal places) confuses the eye of the reader and reduces impact.

Figure 2: Provider education pie chart
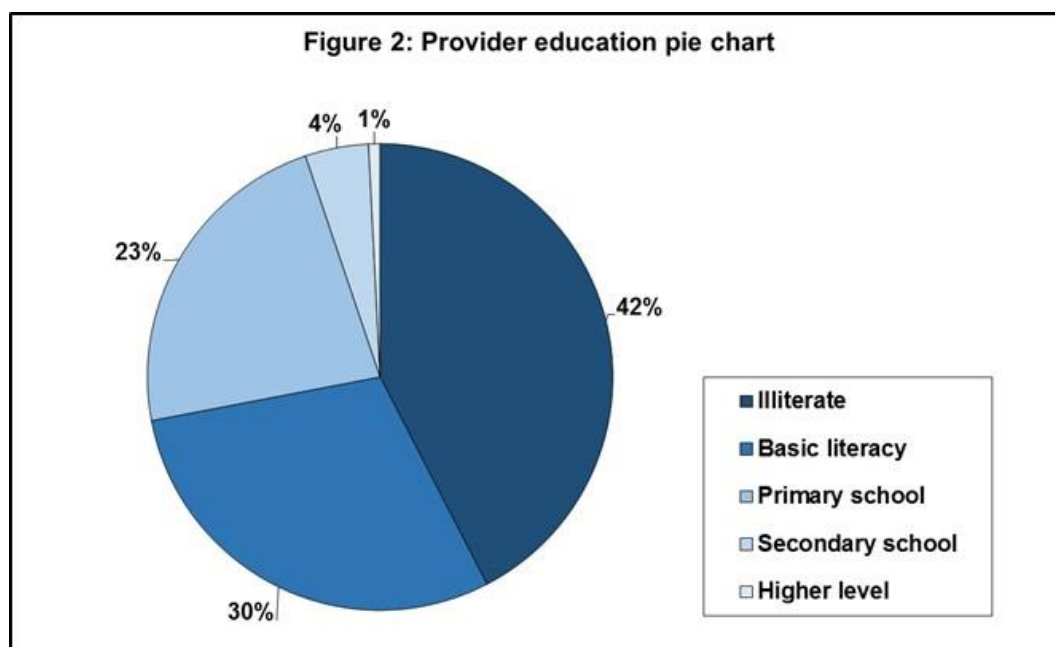
Defining groups for frequency distributions
A key decision in constructing a frequency distribution relates to the choice of groups. In the above examples, the educational attainment groups were predefined. However, we often have to decide how to specify such groups in order to best summarise a given data set. For example,

incomes will need to be grouped into 'income bands' and age data into 'age bands'. The way in which this is done will depend on the aims of the analysis. Demographic analysis, for example, will often aggregate ages into fixed five- or ten-year age bands, such as 0-4, 5-9, 10-14, etc., with a final open-ended group such as '75 and over' (note that these classes are defined such that there is no overlap – the second, for example, relates to 'all children of five years or older but under ten years'). An educationalist, on the other hand might use groups such as 0-5, 6-12, 13-15, 16-18, 19+, where the groups are defined in line with the official age ranges for specific levels of education, for example pre-school, primary, lower secondary, etc.

Just as above, we can construct a frequency distribution based on these groups, showing the number of people falling into each age band. However, here *the definition of groups does involve a loss of information*. Given the number of children in the 5-9 age band, we cannot deduce the ages of the individual children in this group. More frustratingly, we cannot, in the above example, derive the frequency distribution preferred by the educationalist if we are presented with that derived by the demographer. This can be a major problem because we often wish to combine distributions from more than one source. For example, we might know the number of children in primary school and wish to express this as a proportion of all children in the 6-12 age band. If we only know the numbers in the 5-9 and 10-14 age bands, we cannot directly calculate the number we require and have to resort to a weighting procedure based on more or less plausible assumptions.

How should groups be defined? In some cases, such as types of facility or staff salary ranges, official definitions may be most appropriate. If such obvious classifications do not exist or do not serve our purposes, we usually try to balance two conflicting objectives – limiting the loss of information (by using a relatively large number of groups) and providing a simple summary (by using a relatively small number of groups). In general, we would also prefer to make all the group intervals of equal width, because this simplifies comparisons between one group and another. In table 4, for example, a much higher percentage of the studied school-age population are in the second age band than are in the third. However, this is obviously at least partly because this group covers a greater range of ages - seven years as compared to three.

Table 4: Percentage distribution by school age-bands

Age band	Percentage
0-5	35
6-12	37
13-15	15
16-18	14
All	100

Note that the column chart below, which is derived from these data, does not reflect the variations in group ranges. The age bands are used simply as labels for the columns, which are all of equal width. It is the height of the column that shows the percentage falling into each age-band.

Figure 3: Percentage distribution over school age-bands
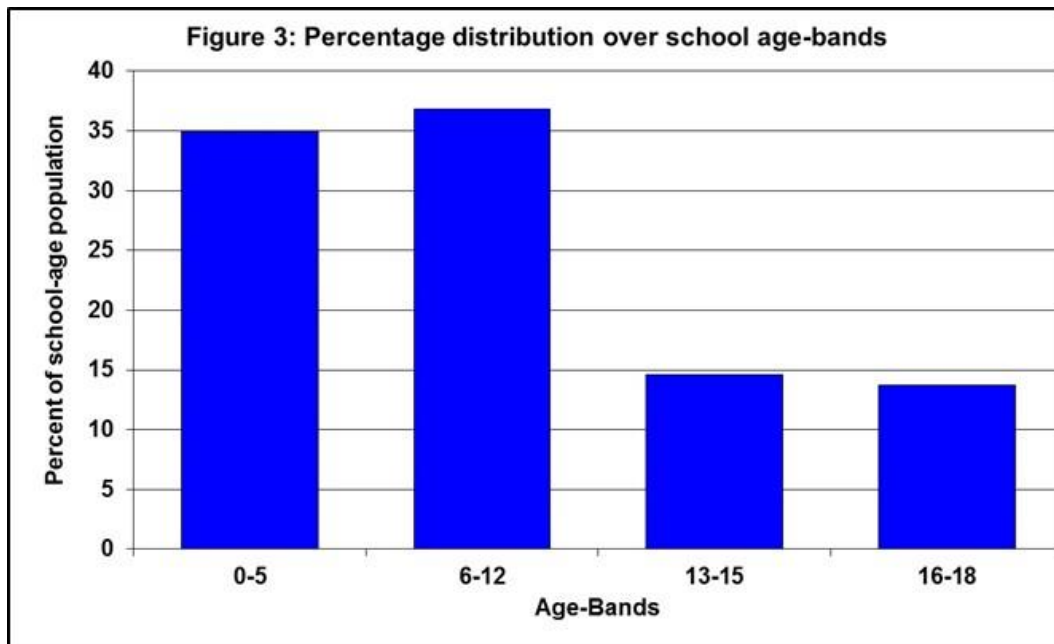
Joint frequency distributions

One of the simplest and yet most powerful techniques for analysing and presenting data involves comparing the frequency distributions of two groups within the study population. Table 5 takes the data used above and disaggregates by the gender of the respondent.

**Table 5: Joint frequency distributions for two or more variables**

Highest level	Men	Women	All
Illiterate	42	64	106
Basic literacy	45	29	74
Primary school certificate	32	25	57
Secondary school certificate	8	3	11
Higher-level qualification	1	1	2
Total	128	122	250

Doing this reveals interesting new information. Although almost the same number of men and women were asked (128 and 122), it would appear from our sample that educational achievement is much higher for the former. We can make the comparison clearer by using the relative frequencies or percentages based on the total number of individuals in each group. Table 6 shows, for example, that 52.5 per cent of women are reported to be illiterate as compared to 32.8 per cent of men. Obviously the conversion to percentages would be even more useful if the numbers in the two groups differed more substantially.

Table 6: Joint distribution using column percentages

Highest level	Men	Women	All
Illiterate	32.8	52.5	42.4
Basic literacy	35.2	23.8	29.6
Primary school certificate	25.0	20.5	22.8
Secondary school certificate	6.3	2.5	4.4
Higher-level qualification	0.8	0.8	0.8
Total	100.0	100.0	100.0

The above table can again be presented graphically in a column chart, as in figure 4.

Figure 4: Bar chart for two variables
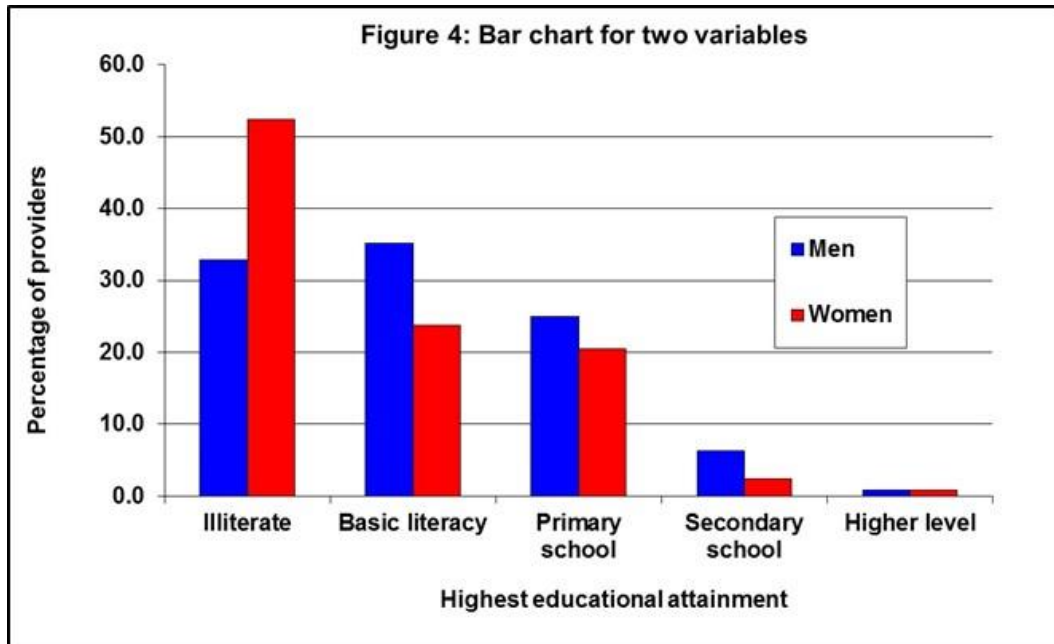
An alternative presentation, that can be useful if we wish to focus on the composition of each class, is obtained by calculating row percentages based on the number of individuals in each education group (table 7).

Table 7: Joint distribution using row percentages

Highest level	Men	Women	All
Illiterate	39.6	60.4	100.0
Basic literacy	60.8	39.2	100.0
Primary school certificate	56.1	43.9	100.0
Secondary school certificate	72.7	27.3	100.0
Higher-level qualification	50.0	50.0	100.0
Total	51.2	48.8	100.0

When interpreting percentage distributions it is always important to check on the absolute size of the denominator on which they are based. For example, the above table shows that 50 per cent of those with a higher-level qualification are men and 50 per cent women. Before getting too excited about this apparent example of gender equality, we should note that only one man and one woman are in this class!

Summary statistics and frequency distributions
Careful examination of the frequency distribution of a variable can be an extremely powerful and robust form of analysis. Unfortunately it is often bypassed. There is a tendency to move too quickly to the calculation of simpler 'summary statistics' that are intended – but often fail – to capture the essential features of the distribution. These usually focus on the derivation of measures:

1. to indicate the overall 'location' of a distribution - how sick, poor, educated is a study population 'on average'?
2. to indicate the extent of 'variation' within that population.

However, the reasons for selecting a particular summary statistic should obviously relate to the purpose for which it is intended. For example, if we ask the question 'Has the recently

implemented intervention reduced the problem of malnutrition among five-year-olds in this village?', there is no doubt as to which of the following possible summary statistics would be more useful:

- Change in mean daily calorie intake of all five-year-olds in the village, or
- Change in proportion of five-year-olds in the village falling below a predetermined minimum calorie requirement.

Bearing in mind the above discussion about the need to present research findings in ways that are appropriate to the various stakeholder groups, appropriate criteria for the selection of summary statistics might be: (1) is the statistic clearly relevant to the specific concern we wish to address; (2) will stakeholders understand how it was derived; and (3) will stakeholders interpret it as intended - that is, are they taking what we would regard as the right message from the information we are providing? We can consider how to apply these criteria by considering some simple examples.

Mean or median?

There is a tendency for quantitative analysis of continuous variables to start by comparing mean values over time - for example by how much has the mean cost of treatment increased - or for different sections of the population, such as how does the mean length of stay in hospital vary between urban and rural populations? The mean is the most commonly used statistic, often seen as the 'natural' measure of central location and used without much thought. However, this is mainly because it is simple to calculate and manipulate. In the days before analysis was done by computer, it was relatively easy to calculate means either by hand or using a calculator. Moreover, given the means for two population groups (for example, two health districts) it was very easy to calculate the mean of the combination as:

combined population mean = $(n_1 \times \text{population mean}_1 + n_2 \times \text{population mean}_2) / (n_1 + n_2)$

Where n_1 and n_2 are the number of observations in the two populations.

On the other hand, we know that most people tend to misinterpret the mean. They assume that it can always be seen as representing the 'typical' value in a population, for example interpreting GDP/capita as the income of a typical person in a given country. In practice this is only a valid interpretation in the case where the underlying frequency distribution is symmetric, for example the so called 'normal' distributions that tend to occur for physical measures such as age-specific heights and weights. For example, in Figure 5 the mean birth weight is 7.5 lbs, which can be seen as a providing a reasonable idea of the typical birth weight of a baby in this population.

Figure 5: 'Normal' distribution of birth weights
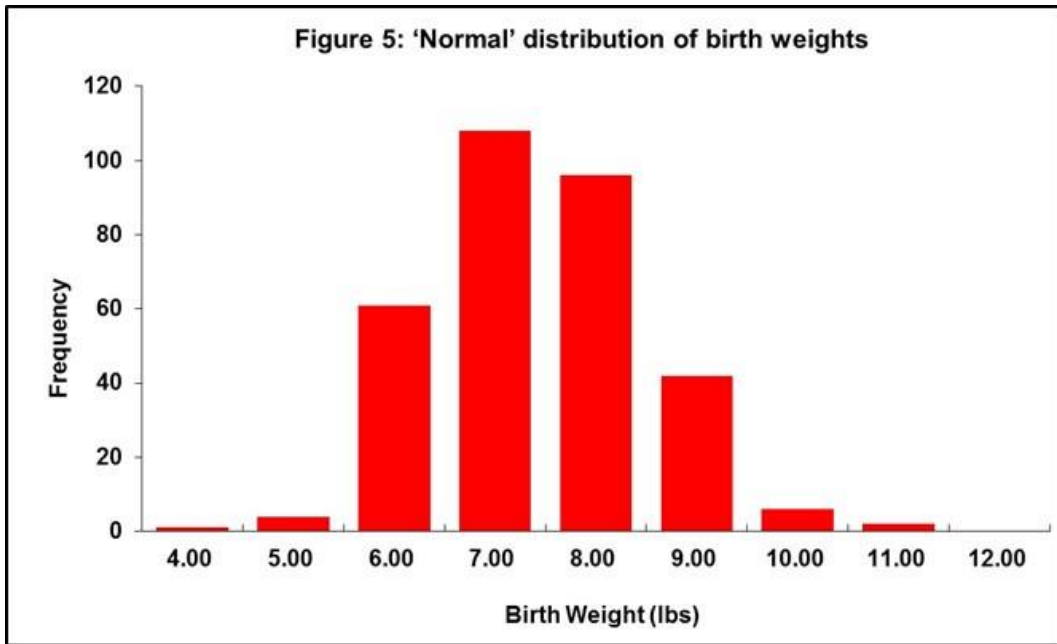
When the distribution is 'skewed', as in Figure 6, the mean can be seriously misleading as an indicator of the situation of most members of the population. It is pulled to the right by the limited number of individuals with high values. Such distributions are very common for variables such as expenditures, income, wealth, lengths of stay in hospital, etc.

Figure 6: Skewed distribution: Mean value difficult to interpret
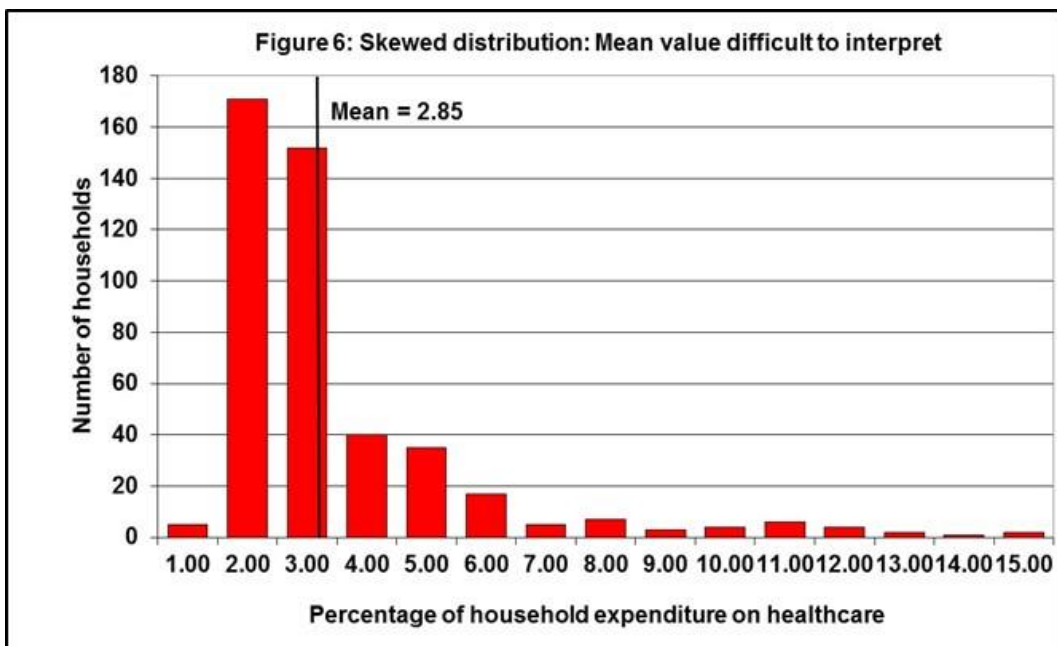
Where the distribution is skewed in this way, the median value may be a better guide. It has the additional advantages of being easy to define and interpret - 'line up the population in order and identify the one in the middle' is relatively easy to explain to all stakeholders. The use of medians may be particularly important in analysis of data sets liable to errors that may include extreme outlier values (it is not unusual, for example, for an individual to accidentally add a zero to a number). Including these outliers in the calculation of the mean, which as indicated above is sensitive to large values, can give rise to biased results. The median is not affected.

An alternative approach sometimes used to deal with outliers is the 'trimmed mean'. For example, a 5 per cent trimmed mean removes the smallest and the largest 5 per cent of data values from the studied population and re-computes the mean using the reduced sample. This can be a useful approach but has the major disadvantage that it often appears somewhat arbitrary and increases the difficulty of explaining results to stakeholders.

Even the median is not much help in more complex distributions, such as the 'bi-modal' in figure 7. This type of distribution is often found where two subgroups are combined, for example patients in urban and rural hospitals. The most useful analysis in such cases involves identifying and separating the subgroups. This again emphasises the need to understand how variables are distributed in order to summarise them in ways that are helpful rather than misleading.

Figure 7: Bi-modal distribution: Both mean and median difficult to interpret
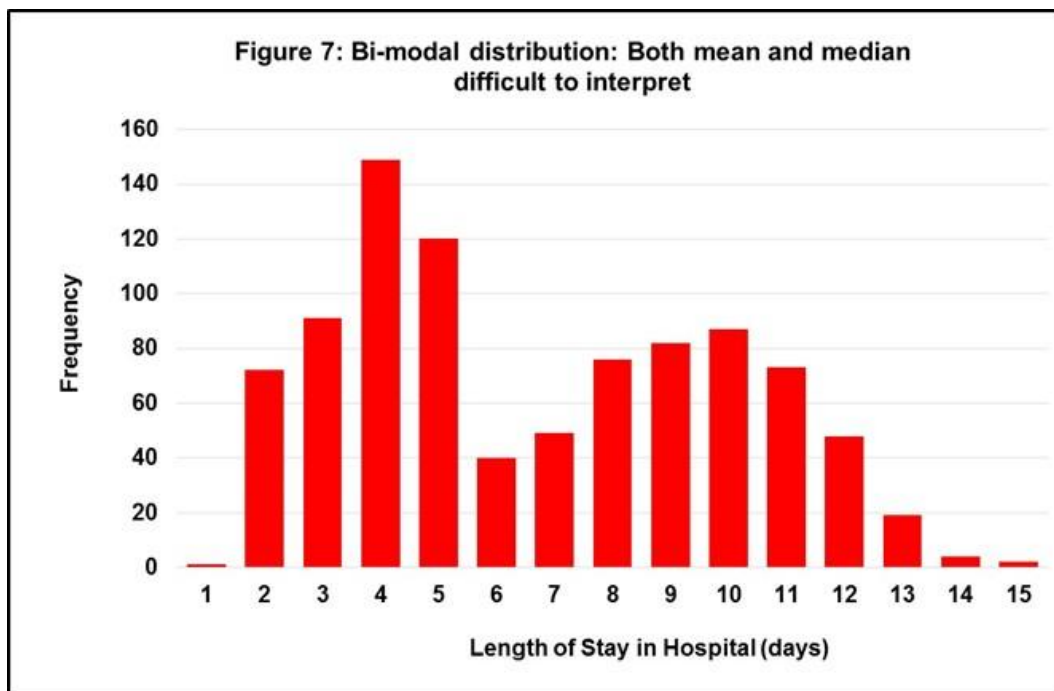
Measures of variation

In a population that has relatively limited *variability* in terms of the *variable* in which we are interested, a measures of location can be seen as reasonably 'representative' of the overall population and there is limited loss of information if we use this as a summary measure. If all those receiving treatment for malaria pay roughly the same amount, we lose little by describing the median or mean payment as 'the cost of malaria treatment'. On the other hand, if the amount paid for treatment of tuberculosis varies substantially across cases, use of the location measure alone would not be an appropriate summary of the data. We would be losing valuable information. Essentially, high variability implies that we have something to explain. Is the variability between urban and rural areas, between facilities within those areas, between patients who are insured and those who are not?

The variance is a measure of variability that considers all the data values relating to a study population. It asked the question 'how far away on average are the data values from the mean'? If we were considering length of stay, for example, and for most patients the stay in hospital was close to the mean, we would say that the distribution was relatively equal – with limited variation 'about the mean'. To calculate the variance we first determine the differences between each value and the mean, the 'deviations from the mean', square each of these differences, find their sum and divide by the number of values:

variance = $\sum (x_i - m)^2 / n$

Note that the size of the variance can often be determined by a limited number of deviations that are much larger than the rest. For example, if we have 100 inpatients and 48 stay in hospital for two days, 50 for three days and two for 20 days, the mean length of stay would be 2.86 days and the variance 6.24. Without the long-stay patients the variance would be 0.25. Simply using the mean and variance to summarise this data would lead to the incorrect interpretation that length of stay varied considerably, while in fact it would be much more useful to report that it appeared to be almost constant but for a few exceptional cases.

This effect results from the squaring of the deviations – squaring a large number produces a very large number. We saw above that the mean was influenced by outliers, but this effect is much more pronounced for the variance. The earliest use of the variance as an indicator of dispersion was in the field of scientific measurement and here it was considered an advantage that it was so influenced by outliers. These were either errors of measurement or extremely interesting data points - both of which required explanation. However, in social research it may often be an undesirable characteristic, first because the errors are typically of less interest (for instance caused simply by poor reporting), and second because it tends to focus analysis on attempts to explain the behaviour or experiences of a small number of individuals in what may be a fairly homogeneous population. Analysis of the differences in length of stay between small rural hospitals and the main teaching hospitals may be interesting, but from a policy perspective it would be probably be differences between the rural hospitals, if they were similar in most other respects, that would be more relevant.

The *standard deviation* is the square root of the variance. It has similar characteristics but also the advantage that it is expressed in the same unit of measurement as the original data. In the example above the standard deviation for all patients would be $\sqrt{6.24}$ = 2.5 days and without the outliers it would be $\sqrt{0.25}$ = 0.5 days. The standard deviation is a very important measure when we consider sampling from a population.

Another commonly used measure derived from the variance is the *coefficient of variation*. This is defined as:

Coefficient of variation = 100 x standard deviation / mean

and provides a measure of *variation relative to the mean*. This is a useful statistic when comparing the variations of data sets that have substantially different means. For example, if we were to compare the variation in incomes for a population of hospital doctors as compared to a population of nurses, we would probably find, if we used any of the measures described above, that the former was considerably larger. However, this would be at least partly due to the generally higher incomes of doctors as compared to nurses. Essentially, the higher the incomes the more scope there is for variation. The coefficient of variation is not affected by this issue. Another advantage is that it is a pure ratio, which has no unit of measurement (because both the numerator and denominator have the same measurement unit). Thus, for example, we could directly compare the variation of incomes that are expressed in different currencies using this measure. It is also unaffected by inflation (as both numerator and denominator are equally affected), so we can consider if income variability has increased over time without worrying about the need to adjust using price indices.

Variances, standard deviations and coefficients of variation are widely used in statistical analysis. As with the mean, this is not because they are always the 'best' measures of variability (they can be easily interpreted for normally distributed variables but not for other distributions) but mainly because they can be readily calculated and manipulated. For example, given the variances of two population subgroups it is easy to combine them to

calculate the overall population variance. However, while they may have technical advantages, all these measures have serious limitations in terms of policy application, given that there is no way to provide a simple explanation of their derivation that would be understood by the great majority of stakeholders.

Alternative, more easily interpreted, measures of variation

Just as the median divides a data set into two halves, with 50 per cent above and 50 per cent below, *quartiles* can be used to divide it into four quarters with 25 per cent of the study population in each. There are three quartile values, usually denoted Q1, Q2 and Q3. If the data are listed in ascending order, Q2 is simply the median. Q1 is the median of the data points below the median and Q3 is the median of the points above the median. A useful and relatively easy to interpret and explain measure of variation is Q3−Q1, the *inter-quartile range*, which includes the 'middle 50 per cent' of a population.

When we have data on a reasonably large population (at least 100 members) we can extend the above to calculate *percentiles*. The p_{th} percentile divides the data into two parts with approximately p per cent having values less than the p_{th} percentile and (100 − p) per cent having values greater. Thus the 50_{th} percentile is the median, the 25_{th} percentile is the first quartile, etc. Other common percentiles, often used with incomes and expenditures, are the 20_{th} (which defines the first 'quintile group') and the 10_{th} (which defines the first 'decile group'). In describing inequality in income of doctors, for example, we might estimate the proportion of total incomes paid to the bottom and top decile groups.

Precise formulae for calculating percentiles are available and used in computer statistical packages. However, because the number of data points is large, an approximation is usually perfectly adequate. For example, if there were 513 data points, it would be reasonable to calculate the quintiles as follows:

$$Q1 \approx 513/5 \approx 103$$
$$Q2 \approx 2 \times 513/5 \approx 205$$
$$Q3 \approx 3 \times 513/5 \approx 308$$
$$Q4 \approx 4 \times 513/5 \approx 410$$

(rounding to the nearest integer) and use these to identify the four quintiles that divide the population into five quintile groups.

Lorenz curves and Gini coefficients

A *Lorenz curve* provides an alternative approach to measuring dispersion based on the cumulative distribution of a variable. The approach is often used for incomes or wealth distribution: for example, 'what share of the total income received by a given population goes to the 20 percent who receive the lowest incomes?', 'what share goes to the lowest 40 percent?', etc. By definition, the shares of each income group will increase as we move up through the income quintiles. However, the approach can also be used to analyse access to services. For example, we can ask 'what percentage of vaccinated children 12−23 months come from the sub-district with the lowest vaccination rate?', 'what percentage from the two sub-districts with the lowest rates?' etc. If we plot those percentages against the total percentage of children 12−23 months, cumulating over sub-districts, we obtain a Lorentz curve illustrating the variation in vaccination rates (figure 8).

Figure 8: Cumulative percentage of children 12-23 months immunised
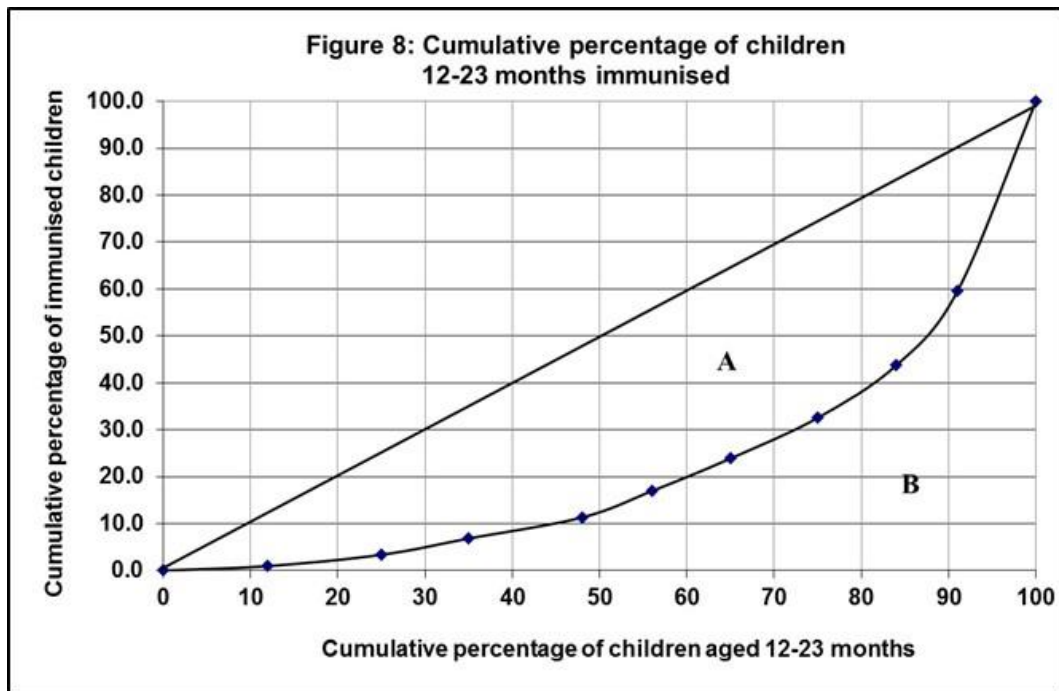
The *Gini coefficient*, the ratio of area A to area (A+B), provides an alternative summary measure of variability that is often used when equity is a priority concern. If there is complete equality, the area A and the Gini coefficient equal 0. As inequality increases, area B becomes smaller and the Gini coefficient approaches 1. For any population the Gini coefficient will lie between 0 (complete equality) and 1 (complete inequality). However, there is no simple interpretation of other coefficient values. It is typically more useful, and certainly easier to communicate with stakeholders, if findings focus on the overall distribution illustrated by the Lorenz curve rather than exclusively on the Gini coefficient.

Concentration curves

Concentration curves (O'Donnell et al. 2008) can be seen as an extension of the Lorenz curve approach to include relationships between two variables. Typically, they show the cumulative percentage of a health status variable plotted against the cumulative percentage of a population ranked by socioeconomic status. For example, figure 9 shows the cumulative percentage of under five deaths plotted against the cumulative percentage of births, ranked by the wealth status of the households in which those births occurred (O'Donnell et al. 2008: Supplementary material). As might be expected the curves both lie above the line of equality because under-five mortality rates decrease with increases in wealth. The fact that the line for India is above that for Mali indicates that inequality in death rates was uniformly higher in India. The interpretation would have been more complicated if the lines for the two countries had crossed at some point. As with the Gini coefficient, it is possible to calculate a *concentration index* if a simple measure of inequality is desired.

Figure 9: Concentration curve
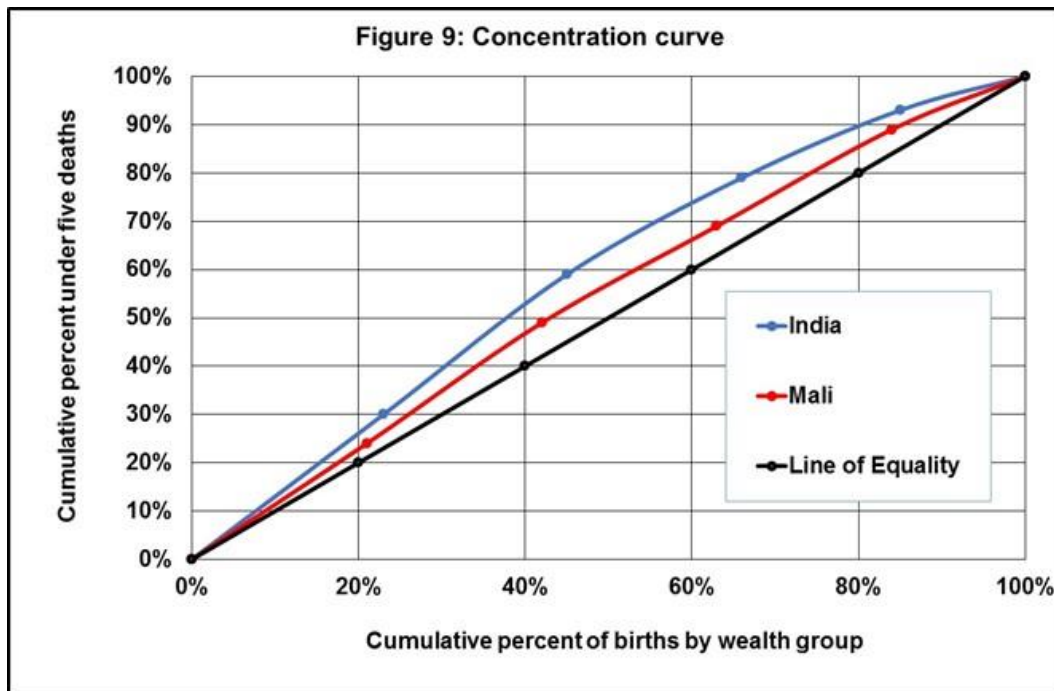
Risk measures: Handle with care

Finally in this section, we can consider measures of 'risk'. These are widely used in health research but again are not well understood by the general population. For example, if the risk of contracting typhoid in an urban area over a one-year period is one in 10,000 and an intervention claimed to have reduced this to one in 20,000, this would probably be reported in local media as 'halving the risk of contracting typhoid'. There might then be a popular call for the intervention to be introduced at scale. However, this would disregard (a) the low risk prior to the intervention and (b) the likely estimation (sampling and non-sampling) errors when attempting to measure such rare events.

As another example, 'risk' and 'odds' are often confused. If we denote the risk of an event as P, then

Risk (P) of an event = number experiencing an event / population at risk.
Relative risk (P(A)/P(B)) = risk in group A / risk in group B.
Odds of an event = number experiencing / number not experiencing = P / (1-P)
Odds ratio = [PA/(1- PA)] / [PB/(1-PB)]

This distinction is particularly important when we consider reductions in risk, which are not equal to reduction in odds, for example:

Risk of malaria before intervention = P(B) = 0.5
Risk of malaria after intervention = P(A) = 0.1
Reduction in risk = 0.1/0.5 = 0.2
Reduction in odds = (0.1 / 0.9) / (0.5 / 0.5) = 0.11

The denominator problem

For the above calculation it is necessary to know the overall size of the population 'at risk'. Similarly, in clinical research one common summary statistic is the proportion or percentage of patients in the intervention and control groups whose condition improves. Calculating such proportions also requires data on the total membership and number improving in each group.

In implementation studies, it is often very difficult to calculate or even reliably estimate these summary statistics because the *denominator* is not reliably known.

For example, we often have only a rough estimate of the number of children who should be immunised or could be sleeping under a net in a given district. Similarly, the catchment population of a facility or actual number of births over a period of time are often unknown. Because of this uncertainty, it is good practice to provide the estimates of both the numerator and denominator alongside any proportion, percentage or risk estimate and to indicate the sources used in the calculation.

4.   Model building

As indicated above, we can regard analysis as essentially concerned with the explanation of variability. For example, why do the costs of care for a given condition vary between patients and/or between facilities? Can this be explained by variations in the severity of the condition or do other factors - patient gender or age, type of hospital, diverse treatment protocols, urban/rural location, etc. - play some role? In general terms, is variation in one variable associated with variation in another and does that association imply some causal relationship? As indicated above, this is an enormous topic to which we can only provide an introduction in this chapter. One excellent online course for those who wish to gain an in-depth knowledge of modelling techniques is that provided by the University of Bristol.

Subgroup analysis

During analysis we will often find that the outcomes of an intervention vary substantially between different subgroups of the target population. It would then seem natural to explore the possibility that the variables that define that subgroup may in some sense have caused the variation or alternatively been caused by it. However, subgroup analysis can be a contentious issue if the subgroups are not predefined. Large data sets containing multiple variables, whether from routine data systems or sample surveys, will often tend to exhibit patterns that may arise purely by chance. The term 'data mining' is often used to describe the process of exploring data sets to discover apparent relationships that may be of interest. It is generally regarded as useful when used to formulate new hypotheses but requiring great caution to avoid being misled - if you search through all possible combinations of variables in a data set containing perhaps 50 or more, there is a high probability that you will stumble across a number of apparent relationships purely by chance.

This is a particular issue in implementation research, where the emphasis is on detailed understanding of processes and an acceptance that relationships between inputs and outcomes may often be mediated by other variables. For example, suppose we find that the prevalence of chronic illness varies by age group and sex as in Table 4.1. If we obtained these findings using a rapid survey as defined above, we must first consider if the sample size was sufficient to provide reasonably reliable estimates of prevalence in each cell of the table.

Table 8: Prevalence of chronic Illness by sex and age group

	Percentage reporting at least one chronic illness	
Age group	Males	Females
15–24	0.55	0.80
25–44	1.79	4.01
45–64	4.91	12.28
65	12.86	20.00
All	1.77	4.25

One of these relationships, between chronic illness and age group, is long-established and well understood - as bodies age they tend to accumulate defects that are linked to various types of chronic illness. The other, the higher prevalence of chronic illness in women in all age groups, is less easily explained. It would not be correct to leap to the conclusion that women are naturally more prone to chronic illness than men. We might consider a range of possible hypotheses exploring, for example, the influence of childbirth, activities mainly undertaken by women and men, whether women were more likely to report illness than men or less likely to receive treatment for acute illnesses that then became chronic conditions. We might be able to examine some of these hypotheses by further analysis of this or other data sources, or by undertaking qualitative studies such as described in Chapters 9 and 10. The key requirement for researchers is to ensure that they have convincing evidence before advancing one or other theories to explain such observations.

## *Controlled and confounding variables*

We sometimes describe such an analysis as one that assesses the relationship between inputs and outcomes *controlling* for other factors. Typically we know that in practice a very large number of other factors may influence this relationship, for example occupation, level of education, socioeconomic status, household size, type of dwelling, rural/urban location, etc. As indicated in Chapter 5, random allocation of subjects to subgroups would allow us to argue that the potentially 'confounding' effects of such variables average out. That will almost never be possible in the type of interventions we are considering and we therefore need to find some way to allow for these effects.

Cross-tabulating by all such factors, even with an apparently large data set, would almost always result in the numbers in most cells being too small to permit analysis. One alternative is to construct a model of the relationship between outcome and inputs that takes into account the effects of other confounding variables. This typically involves very strong assumptions both as to the nature of the multiple relationships between these variables and their individual distributions – assumptions that are often not adequately recognised or tested. As discussed above, it can be argued that the explicit intention to change implementation practice and influence a wide range of stakeholders requires implementation researchers to set higher standards than those conducting more exploratory research.

Models and presentation of findings

Models are typically very simplified versions of reality and we should be very cautious in their interpretation. In particular we should recognise that most stakeholders will typically have little understanding of the assumptions underlying those models. Modelling may be useful to explore our data but should be seen as an intermediary stage in the generation of findings that can be readily comprehended and interpreted. As with the step from distributions to summary measures, we should proceed cautiously and try to ensure that we understand the underlying form of the relations that we are trying to model.

Just as we can understand a great deal about individual variables by examining frequency distribution, much can be learned about two-way relationships from simple scatter diagrams. Figure 10 illustrates that such relationships can take a great variety of forms. It shows possible scatter diagrams of a sample of out-of-pocket payments for inpatient care plotted against length of stay in four hypothetical hospitals, with very different policies on fees and with patients covered by different types of health insurance. Only the first might reasonably be assumed to reflect a linear relationship.

Figure 10: Scatter diagrams illustrating different types of relationship
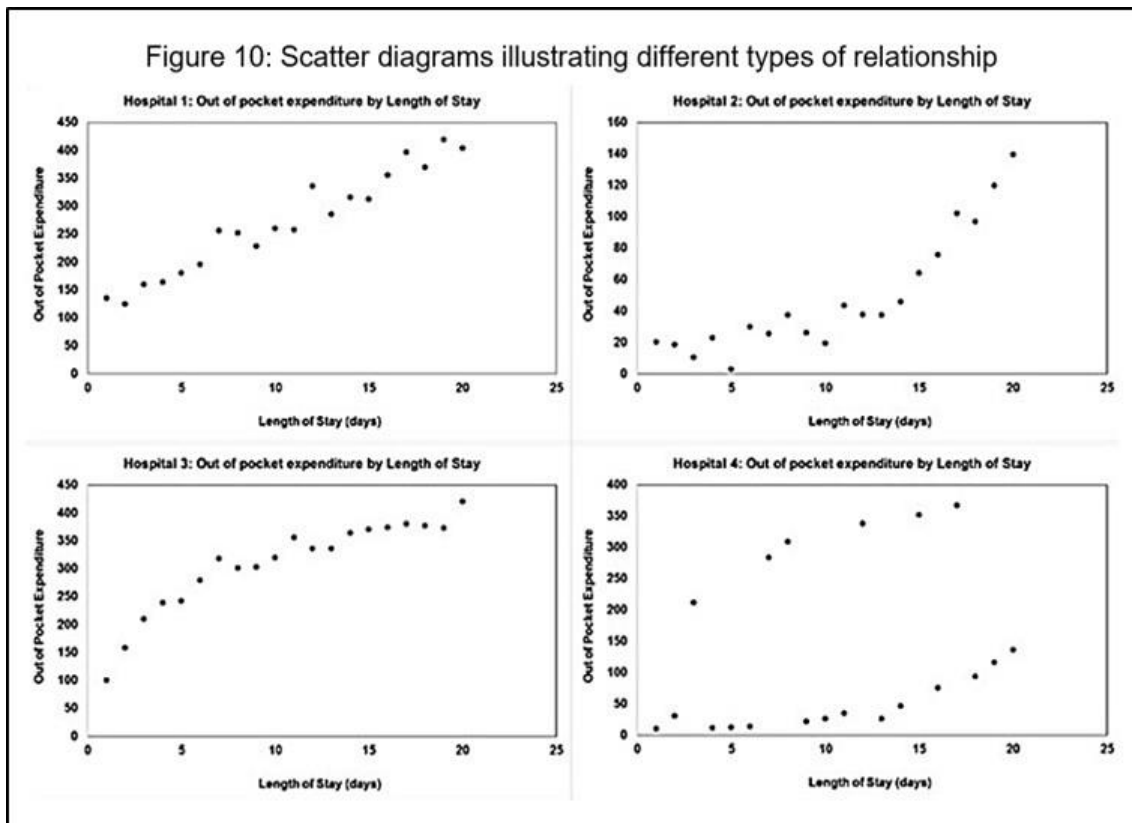
Our common, often unspoken, assumption of linear relationships between variables is frequently not only incorrect but may in many instances be simply irrational. The number of cases of tuberculosis identified cannot increase linearly with expenditure on case finding, because finding cases will become increasingly more difficult once the 'low-hanging fruit' have been identified. Hospital net revenues cannot increase linearly with the number of inpatients, because the marginal cost of an inpatient will decline as the number increases.

The linear regression model
By far the most common approach to model building is the use of some form of linear model and we can use this to illustrate modelling possibilities and limitations. The simple linear regression model is illustrated in figure 11. It is usually expressed by an equation of the form:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Where:

y_i is the value of a response (outcome) variable for the ith observation.
x_i is the value of an explanatory (input) variable for the ith observation.
ε_i is the value of a random error term for the ith observation.

This model is the equation of a straight line where:

α is the intercept (predicted value of y when x=0) and
β is the slope.

Figure 11: The linear regression model
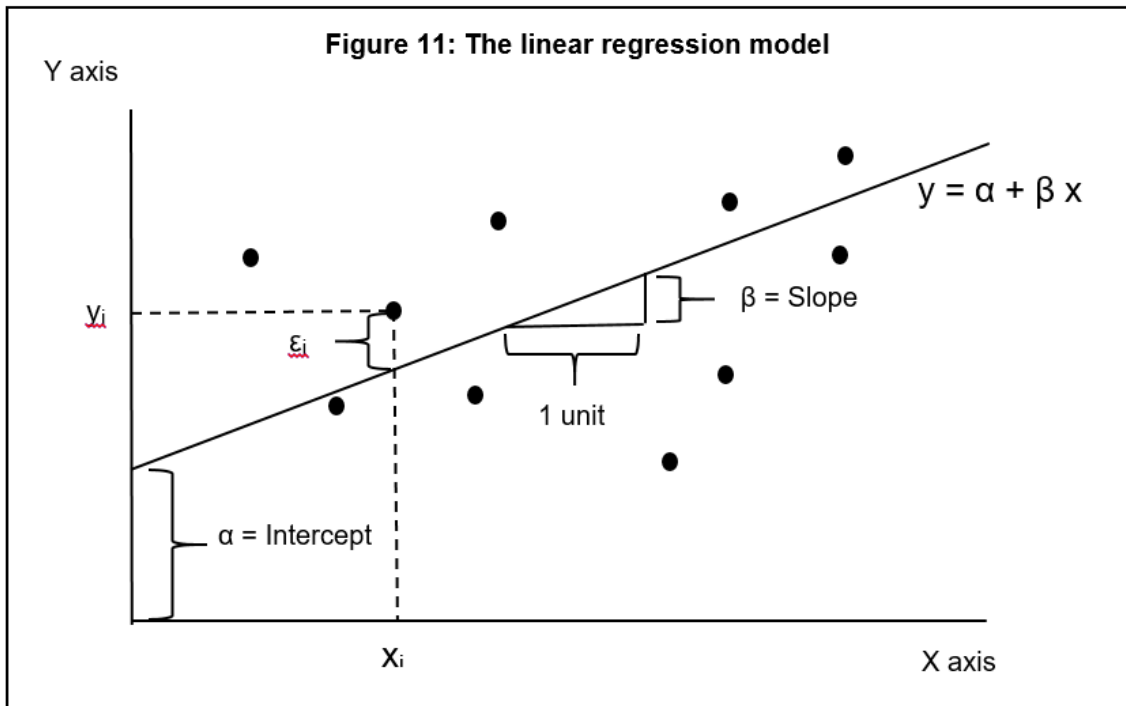
Regression assumptions

The following strong assumptions, which many researchers choose to ignore, are required in order to argue that a regression model is appropriate:

- The relationship between X and Y is linear;
- The values of the independent variable X are assumed fixed (not random) - the only randomness in the values of Y comes from the error term ε;
- The errors ε_i are uncorrelated (independent) in successive observations;
- The errors ε_i are normally distributed with mean 0, variance σ^2 [$\varepsilon \sim N(0, \sigma^2)$].

We choose α and β such that the sum of squares of deviations from the regression line [Σ(observed value of y_i at x_i - predicted value of y_i at $x_i)^2$] is minimised. This is known as the error sum of squares (ESS) about the regression. ESS/(n-2), where n is the number of observations, provides an unbiased estimate of σ^2.

Variance components and the coefficient of determination

The error sum of squares can be compared to the sum of squared deviations about the mean (TSS) to see how much of this can be 'explained' by fitting the regression line. The division of the sum of squares about the mean (TSS) into two 'components', a regression sum of squares (RSS) and an error or residual sum of squares (ESS), is the simplest example of the 'variance components' approach to model building, which plays a central role in multilevel modelling. If we write $\hat{y}_i$ for the value of y_i predicted by the regression equation and $\overline{y}$ for the mean value of the observations, we can express the deviation of y_i from $\overline{y}$ as the sum of the deviation of y_i from $\hat{y}_l$ plus the deviation of $\hat{y}_i$ from $\overline{y}$:

$$(y_i - \overline{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})$$

If we square both sides and sum over all values of y_i, we can derive the following result:

$$\text{TSS} \left[\Sigma(y - \overline{y})^2 \right] = \text{ESS} \left[\Sigma (y - \hat{y})^2 \right] + \text{RSS} \left[\Sigma(\hat{y} - \overline{y})^2 \right]$$

The ratio R^2 = RSS/TSS = 1 - ESS/TSS is known as the *coefficient of determination* and is often loosely described as the proportion of variance 'explained' by the model.

Residuals

The use of the phrase 'explained by the regression line' should not be taken literally. It refers simply to the above ratio, which is interesting only if all the assumptions made in defining our model are correct - this is rarely the case. One way of exploring the value of our model is to look at the deviations of observations from the value 'predicted' by our regression line - the 'residuals'. We do this using a scatter plot with the explanatory variable (X) on the horizontal axis and the residuals on the vertical axis as in figure 12.

Figure 12: Residual plots
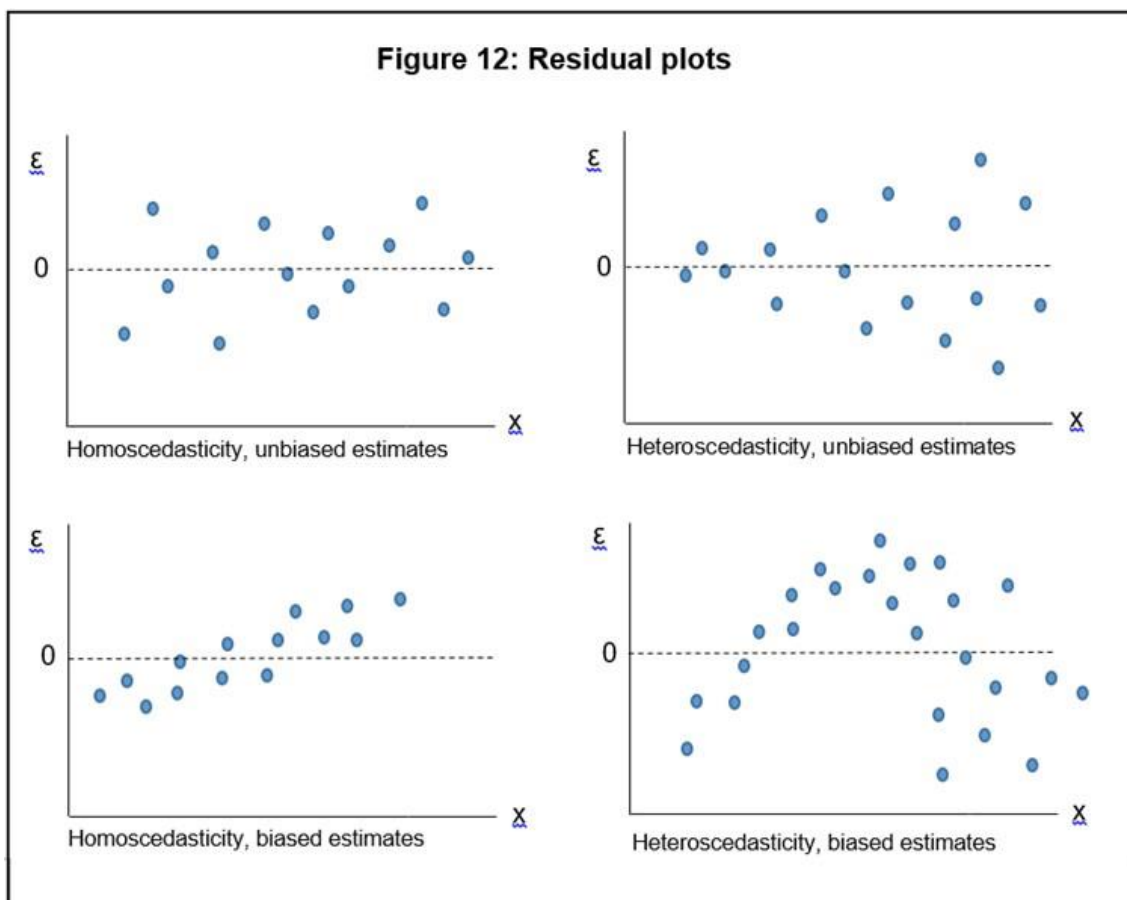
Specification errors

As indicated above, we use models to allow for the effects of a variety of potentially confounding variables. To do this we construct a multiple regression model:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \delta_1 z_1 + \delta_2 z_2 + \ldots + \varepsilon$$

where:

y	is the response variable
x_i	are known explanatory variables
z_i	are known confounding variables

However, one often intractable issue is that there are typically a range of factors which we have either ignored or cannot measure. The true model should be written:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \delta_1 z_1 + \delta_2 z_2 + \ldots + \gamma_1 c_1 + \gamma_2 c_2 \ldots + \varepsilon$$

where*:*

y	is the response variable
x_i	are known explanatory variables
z_i	are known confounding variables
c_i	are unknown confounding variables
ε	is an error term*.*

This is described as a 'specification' error. In general, omitting such variables from the model has serious implications in terms of undermining the basic assumptions identified above.

Statistical inference in regression models

As discussed above, with the widespread availability of statistical software, it is expected that, where data have been collected using probability sampling, all estimates will be accompanied by estimated error margins. For example, in the simplest case of a random sample of size n we know that we can estimate 95 per cent confidence limits for a population mean as:

sample mean $\pm$ 2 s/$\sqrt{n}$

Where the term s/$\sqrt{n}$ is the 'standard error' of estimation (s.e.). It was also indicated above that for other probability sampling designs the formula for the standard error will vary, but the formula remains the same and can be extended to other statistics:

estimated value $\pm$ 2 s.e.

Multilevel modelling

Given that regression estimates will also require to be accompanied by error margins, we again have to address the issue that most surveys will use a sample design that involves cluster sampling at one or more levels. For example the DHS surveys typically involve:

- stratification by states/provinces and then by urban and rural areas;
- a PPS cluster sample of enumeration areas within each stratum;
- a systematic sample of 30 households per cluster.

As discussed above, failure to allow for the larger sampling errors associated with cluster sampling can result in the confidence limits for estimates that are too narrow, and incorrect assessment of tests of statistical significance. With a cluster sample the error sum of squares (ESS) has two components:

ESS (about mean) = ESS (between clusters) + ESS (within clusters)

If the random sample formula for the error sum of squares is used, estimates of model parameters may be unbiased but estimated confidence limits will typically be far too narrow - that is, we will be substantially exaggerating the precision of our estimates. Multilevel modelling (Rashbash et al 2012, Diez Roux 2009, Goldstein 1999) explicitly builds the variation between clusters into the model and estimates the between-cluster variation.

Random intercept model

We can allow for between-cluster variation by formulating the model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + [u_j + \varepsilon_{ij}]$$

Where:

y_{ij} is the value of y for individual i in cluster j
x_{ij} is the value of x for individual I in cluster j
u_j is the deviation of the mean of cluster i from the global mean

We then have two random variables in the equation:

$u_j \sim N(0, \sigma_u)$
$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon)$

And can obtain correct estimates for: β_0, β_1, σ_u^2, σ_ε^2

Note that sometimes the variation between clusters may itself be of interest. For example, if we have clustered by health facility, we can estimate the proportion of total variability 'explained' by between-facility variation (Lopez-Cevallos and Chi 2009).

Sample survey software

Multilevel modelling can similarly be used in a wide range of other contexts where relationships exist between different 'levels' of a health system (district, facility, doctor, patient, etc.). Using modern survey analysis software it is relatively straightforward to describe even complex sampling design and obtain appropriate parameter estimates. These packages include the 'usual suspects': SAS, STATA, SPSS and some more specialist software such as MLwiN, etc. They can all readily address the most common health survey designs involving cluster sampling and unequal sampling probabilities.

Example: A multilevel analysis of self-reported tuberculosis disease in a nationally representative sample of South Africans was undertaken based on the 1998 DHS (Harling et al. 2008, Harling 2006)). Individual and household-level demographic, behavioural and socioeconomic risk factors were taken from the DHS; data on community-level socioeconomic status (including measures of absolute wealth and income inequality) were derived from the 1996 national census.

Of the 13,043 DHS respondents, 0.5 per cent reported having been diagnosed with tuberculosis disease in the past 12 months and 2.8 per cent reported having been diagnosed with tuberculosis disease in their lifetime. In a multivariate model adjusting for demographic and behavioural risk factors, tuberculosis diagnosis was associated with cigarette smoking, alcohol consumption and low body mass index, as well as a lower level of personal education, unemployment and lower household wealth. In a model including individual- and household-level risk factors, high levels of community income inequality were independently associated with increased prevalence of tuberculosis.

The multilevel analytic approach was seen as allowing for the differentiation between community- and individual-level mechanisms in the relationship between socioeconomic status and tuberculosis. Furthermore, these data allow strong inferences to be drawn regarding risk factors for tuberculosis disease across the country: a nationally representative cross-sectional survey provided evidence on individual and household characteristics, while South African census data provided robust estimates of the true community-level socioeconomic characteristics across the nation

References

Bennett, S.; Woods, T.; Liyanage, W.M. and Smith D.B. (1991) 'A Simplified General Method For Cluster-Sample Surveys of Health in Developing Countries', *World Health Statistics Quarterly* 44: 98−106, http://apps.who.int/iris/bitstream/10665/47585/1/WHSQ_1991_44%283%29_98-106_eng.pdf?ua=1 (accessed 30 March 2015)

Diez Roux, A. V. (2009). Next steps in understanding the multilevel determinants of health. Journal of Epidemiology and Community Health 2008(62):957-959. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/64000/diez%20roux_next%20steps%20in%20understanding%20the%20multilevel%20determinants%20of%20health.pdf;sequence=1

Eurostat (2007) Handbook on Data Quality Assessment Methods and Tools. Wiesbaden: European Commission, http://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK ON DATA QUALITY ASSESSMENT METHODS AND TOOLS_I.pdf (accessed 30 March 2015).

Goldstein, H. (1999) *Multi-level Statistical Models*. London: Arnold. http://www.ats.ucla.edu/stat/examples/msm_goldstein/goldstein.pdf

Goodson, J.L., Kulkarni, M.A., Eng, J.L.V., Wannemuehler, K.A., Cotte, A.H., Desrochers, R.E., Randriamanalina, B. and Luman, E.T. (2012). Improved Equity in Measles Vaccination from Integrating Insecticide-treated Bednets in a Vaccination Campaign, Madagascar. *Tropical Medicine and International Health* 17(4): 430−37

Harling, G., Ehrlich, R. and Myer, L. (2008). The Social Epidemiology of Tuberculosis in South Africa: A Multilevel Analysis. *Social Science and Medicine* 66: 492−505. https://www.researchgate.net/publication/5923792_The_social_epidemiology_of_tuberculosis_in_South_Africa_A_multilevel_analysis

Harling, G. The Social Epidemiology of Tuberculosis in South Africa: A Multilevel Analysis. Thesis. Master of Public Health. University of Cape Town. https://open.uct.ac.za/bitstream/handle/11427/9326/thesis_hsf_2006_harling_g.pdf?sequence=1

Lemeshow, S. and Robinson, D. (1985). Surveys to Measure Programme Coverage and Impact: A Review of the Methodology Used by the Expanded Programme on Immunization. *World Health Statistics Quarterly* 38.1: 65-75

Lin, Mingfeng, Henry C. Lucas Jr, Galit Shmueli (2013). Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. Information Systems Research 24(4):906-917. http://eprints.exchange.isb.edu/280/1/isre%252E2013%252E0480.pdf

Lopez-Cevallos, D.F. and Chi, C. (2009). Health Care Utilization in Ecuador: A Multilevel Analysis of Socio-Economic Determinants and Inequality Issues. *Health Policy and Planning* 25(3):209−18. http://heapol.oxfordjournals.org/content/25/3/209.full.pdf+html

Myatt, M.; Feleke, T.; Sadler, K. and Collins, S.(2005) 'Field Trial of a Survey Method for Estimating the Coverage of Selective Feeding Programmes', *Bulletin of the World*

Health Organisation 83(1):20–6. www.who.int/bulletin/volumes/83/1/20arabic.pdf (accessed 30 March 2015)

O'Donnell, O.; van Doorslaer, E.; Wagstaff, A. and Lindelow, M. (2008). *Analyzing Health Equity Using Household Survey Data*. Washington, DC: The World Bank. http://siteresources.worldbank.org/INTPAH/Resources/Publications/459843-1195594469249/HealthEquityFINAL.pdf (accessed 30 March 2016)

Prudhon, C. and Spiegel, P.B. (2007). A Review of Methodology and Analysis of Nutrition and Mortality Surveys Conducted in Humanitarian Emergencies from October 1993 to April 2004. *Emerging Themes in Epidemiology*, www.ete-online.com/content/pdf/1742-7622-4-10.pdf (accessed 30 March 2015)

Rashbash, Jon, Fiona Steele, William J. Browne and Harvey Goldstein (2012). *A User's Guide to MLwiN*. Centre for Multilevel Modelling, University of Bristol. http://www.bris.ac.uk/media-library/sites/cmm/migrated/documents/manual-web.pdf

SMART (2005). *Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: SMART Methodology*. The SMART Initiative, www.smartindicators.org/SMART_Protocol_01-27-05.pdf (accessed 30 March 2015)

Tipping, Jill and Malcolm Segall. (1996) How to do (or not to do) . . . Using a longitudinal illness record to study household health care decision-making in rural communes of Viet Nam. *Health Policy and Planning* 11(2): 206-211. http://heapol.oxfordjournals.org/content/11/2/206.full.pdf

Turner, A.G.; Magnani, R.J. and Shuaib, M. (1996). A Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design. *International Journal of Epidemiology* 25.1: 198–203

Walker, N.; Bryce, J. and Black, R.E. (2007). Interpreting Health Statistics for Policymaking: the Story Behind the Headlines. *Lancet* 369: 956–63

Chapter 9: Qualitative methods

Donald Skinner
Stellenbosch University

1.   Introduction

Qualitative methodology has always been part of the set of tools used by health researchers (Mack et al. 2005). It has become more important now with the rise in lifestyle illnesses and epidemics where human behavioural choices influence risk (Daniels et al. 2016). There are many pressing research questions about health behaviour and risk-taking that cannot be answered with numbers − for example, why have some AIDS interventions not worked as expected because not all those offered medication were willing to accept it. While research into public health and health systems has been predominantly quantitative in nature, qualitative methods can make a valuable contribution, in particular by allowing an increased understanding of the perceptions, motivations and behaviours of those involved. This approach can be used to generate detailed, in-depth knowledge, often by employing what might appear to be relatively simple techniques of communication and observation, but which require considerable expertise to use effectively (Serekoane et al. 2014). A knowledge of, and experience with, qualitative methods is therefore an essential addition to the set of skills required for implementation research (Atkins et al. 2015).

In this chapter qualitative methodology is described in the context of the interpretive research paradigm. Schwartz-Shea and Yanow (2012) explain the underlying philosophy of this approach as follows:

> *"A researcher can interview based on the belief that she is going to be able to establish 'what really happened' in a setting. This reflects a realist - objectivist methodology … . Or a researcher can interview based on the belief that there are multiple perceived and/or experienced social 'realities' concerning what happened, rather than a singular 'truth'. In this view, the researcher would assume that event narratives are likely to vary depending on the perspective (political, cultural, experiential, etc.) of the persons being interviewed. This approach reflects a constructivist - interpretivist methodology that rests on a belief in the existence of (potentially) multiple, intersubjectively constructed "truths" about social, political, cultural, and other human events; and on the belief that these understandings can only be accessed, or co-generated, through interactions between researcher and researched as they seek to interpret those events and make those interpretations legible to each other."* (p4)

This interpretive process is fundamental to qualitative research, in that it allows analysis to proceed beyond the purely descriptive. Interpretative analysis has also been of vital importance in world history, as it has allowed people to move beyond the obvious and immediate constraints of their environment (Fossey et al. 2002). Examples include Steve Biko and Martin Luther King, in their arguments against racism, and the suffragettes, who criticised sexism and patriarchy. These people looked at their context, reviewed their own situation in relation to that context, made interpretations and came to conclusions that guided their future actions. Good qualitative research draws on this capacity for interpretation and applies *critical* and *scientific* approaches (Cooper and Endacott 2007) developing systems of enquiry that are *explicit* and *systematic*.

Uses of qualitative research
Qualitative research can be valuable in many applications:

- In the early phase of a study to explore an area on first entry into a field, to refine concepts or to formulate or clarify hypotheses. For example, in a recent project researchers and activists could not agree as to what constituted an OVC (orphan or vulnerable child). It was seen as important to seek a community perspective on this concept at a very early stage of the research ([Skinner et al. 2006](#)).
- Alongside other types of research in order to gain an additional perspective on a problem. For example, in a project looking at the risks for women in *shebeens*, unlicensed bar or drinking clubs, it was agreed that while the risks were generally well known, an in-depth understanding of the perspectives and attitudes of the various stakeholders was required to develop interventions. Qualitative methods were used to gain this understanding ([Sikkema et al. 2011](#), [Watt et al. 2012](#)).
- To clarify unexpected or significant findings from quantitative studies. For example a number of such studies have revealed differences between men and women as to the meaning of the term 'concurrent partners' and discordant reports as to the numbers of partners individual have had ([Nnko et al. 2004](#)).
- When the aim is to get an in-depth sense of what people think of a particular object, event or construct. For example, it was central to an evaluation of the role and implementation of the South Africa Truth and Reconciliation Commission ([Skinner 2000](#)); to gaining a real understanding of stigma ([Chambers et al. 2015](#)); and the cultural impact of male medical circumcision ([Lukobo and Bailey 2007](#)).
- In standalone research on a difficult topic or with a hard-to-reach population. For example, research on sex workers or with drug users ([Watt et al. 2016](#)).
- In process evaluations, for example to observe workshop interventions and assess how they are received.
- To provide evidence in support of a social action model (for example, [Freire 1970](#)).
- Every day by anyone to examine their own context. What do I really want to achieve by undertaking this research? What do my colleagues think of me? What do I need to do to advance my career?

Theory of qualitative research

Different models exist to define what is meant by 'science' and the development of knowledge over time. The classical model, as used in physics for example, is that of 'falsification', the process by which incorrect theories are rejected on the basis of empirical evidence. Medical science tends to follow an alternative 'hypothesis-induction' approach – a circular process of reasoning in which hypotheses are specified, for example that a given drug will improve the condition of patients with a given disease, and then repeatedly tested, perhaps with variations, for example in terms of the dosage used, over a series of 'trials'. Typically, in any given trial some patients will respond well, some less well, some not at all and possible some badly. Over time there will be a judgement, based on a statistical analysis of all those trials considered to have been appropriately undertaken, as to whether there is sufficient evidence to promote the treatment, possibly in a modified form and possibly only for some categories of patient. A range of methods have been developed to improve the rigour of both observational and experimental trials. The Double-Blinded Randomised Controlled Trial ([Sibbald and Roland 1998](#)) is considered the 'gold-standard', and there have been suggestions that it should be applied beyond the medical field ([May 2012](#)).

The social sciences offer a different approach which led to the development of, amongst other approaches to knowledge, qualitative research methodologies. First we need to look at what is different about the social world and why we need such a different approach. If you do an experiment by mixing two chemicals in controlled conditions they should always react in the same way. Likewise, if someone has an infection and a proven treatment is applied, the person has a high probability of being cured. If they are not, then the reasons can often be clearly determined, including incorrect dosage or the presence of resistance.

In the social world there is generally much less predictability. For example, consider the responses to an anti-smoking campaign. The information provided may be very clear but the responses will typically vary widely. Explanations for these different responses will vary even more widely, between public health officials, different community members, smokers and non-smokers. Even a single individual may offer different explanations, depending on the context in which they are asked. Research within the social world has to cope with these variations. Multiple theories have been developed around behaviour and what influences it:

- Theory of reasoned action (Albarracin et al. 2001);
- Health beliefs model (NIH undated);
- Information motivation behaviour model (Chang et al. 2014);
- Lay beliefs theory (PHAST 2011);
- Social representations (Howarth et al 2004).

One particularly important approach involves looking at behaviours and decisions in relation to the context within which they are located. It is evident that context strongly influences how people think and act. Not a simple matter, those contextual influences may include culture, language, access to resources, gender, education, age, time, date, knowledge of health issues, social norms, and multiple other factors. The research task is therefore to understand the social world and the context within which individuals are situated in order to make sense of their different responses to external stimulate, such as health system interventions. While the traditional world of scientific experimentation assumes a single empirical reality, the social world can only be examined by reference to the contexts in which that reality is known, and through all the filters that individuals use to understand and respond to that world. Our basic assumptions can therefore be expressed as:

- Philosophies of human behaviours and beliefs direct the qualitative research approach.
- Those behaviours and beliefs are shaped by the context within which an individual is situated, and in turn influence that context.
- Each behaviour and belief carries meaning.
- These systems of social and personal meanings cannot be adequately described by any statistical model.
- Qualitative research should take an interpretive and subjective approach.
- Researchers are part of the research and must examine the ways in which their own beliefs and context will influence their interpretations of the meanings of others.

Qualitative research attempts to understand people and events by acknowledging the full complexity of their contexts and interpreting behaviours and beliefs within those contexts.

> *"Qualitative research … consists of a set of interpretive, material practices that makes the world visible. These practices ... turn the world into a series of representations including field notes, interviews, conversations, photographs, recordings and memos to the self. At this level, qualitative research involves an interpretive, naturalistic approach to the world. This means that qualitative researchers study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them.* (Denzin and Lincoln 2000:3)

Paradigms and theories

A *paradigm* is the collective understanding that we have of our world, including our culture, ideology, assumptions of power, and perceptions of ourselves in that world.

Theories are statements about rules and systems that direct how events happen in the world. These include physical, chemical and biological processes, but here we focus on human

behaviour. There are 'grand' theories, for example relating to the importance of self-efficacy in achieving health goals ([Strecher et al. 1986](#)), or provider and consumer behaviour given the relationship between supply and demand in a market. However, in general, qualitative researchers are primarily concerned with the theories they can construct based on their data, interviews and observations. These are normally only directly applicable to the research subjects but with care inferences can often be made about the broader world.

Context

Many theorists have spoken about the importance of context in shaping our behaviour. In different contexts, data themselves can take on new meanings. Traditionally, 'data' are characterised as individual pieces of information, for example as collected in a spreadsheet, each divorced from its context. However, data in qualitative research remain integrated as far as possible in the context from which they are drawn. Even the research process is essentially part of that context.

What does context mean? It may include environmental, physical, political, social, religious, ideological, cultural, and economic factors because qualitative research takes place in a natural rather than an experimental setting. It is clearly not possible to take everything into account, but researchers should try to maintain as much of what is important as possible. Routinely comparing notes amongst members of a research team will enhance the process of identifying those aspects of context that are likely to influence results. Ideally you collect data while interrupting the existing context as little as possible.

Subjectivity

Subjectivity is central to qualitative research. It is perceptions that are important. Subjective knowledge is ultimate based on the testimony of our senses - everything is understood and interpreted through the eyes, ears and brains of individuals. The aim of qualitative research is to fully appreciate the perspective of the respondents, which means removing yours as far as possible from the interaction. Bear in mind that two respondents in the same situation may have very different perceptions of what is happening – consider for instance the complex exchanges and interpretations of those exchanges that may take place between a man and a woman who meet for the first time in a student bar. You have to make sense of such complexities, bearing in mind that you can never get access to the total reality of any respondent.

Everything that we as researchers perceive is affected by who we are and our context. Likewise the information we are given will be influenced by that context, even who we are able to talk to and what we may observe. Compare the different interactions and interviews that might be undertaken in a South African township by a white male academic and a black female local community member. These realities have to be factored in at all times during the research process, when developing the protocol, preparing to entering the field, interviewing and observing, doing the analysis and interpretation and writing reports.

Meaning and interpretation

Pursuit of meaning is the core focus of qualitative research. What lies behind observed behaviours and spoken or written words? What do respondents mean when they use terms such as vulnerability, community, disability or mental illness? Rather than counting and measuring, the core task is to translate these meanings into a report so that the understanding can be shared. Remember that meanings may differ between people even if they have apparent common contexts.

Interpretation is the process of finding meaning. Analysis in qualitative research involves looking at what is shown in our data, understanding it and then offering interpretations within a given context. We explore the meanings of concepts – vulnerability, community, disability or

mental illness - and construct 'working hypotheses' relating to those concepts as we go through the data, constantly revisiting and refining them.

Reflexivity

In most research activities we try to eliminate the influence of the researcher. In questionnaire surveys, for example, we train enumerators to ask each question using precisely the same words and we encourage them to adopt a formal approach to each interview, trying not to let their own personality influence the answers provided in any way. In qualitative research, you are the research instrument. The aim is to fully acknowledge and understand the influence that you are having on each interaction. Reflexivity, the process of continuously assessing how your actions, values and perception impact on each stage of your research has been described as:

> "*a strategy that researchers can use for the purpose of understanding the phenomenon under exploration and accurately portraying the meaning made by participants and where self-examination allows assumptions and biases that could affect the study to be understood*" ([Lambert et al. 2010:321](#))

Genuine reflexivity is a skill that develops slowly through experience and continues to develop throughout life. It involves learning how to *really* listen and observe, a skill useful not only in research but in life. It requires a capacity for self-examination and an appreciation of the role of your own context and subjectivity. You need to learn how to analyse your perceptions and how they impact on your responses. There is an inevitable conflict between having to be present as yourself, and aware of how you are perceived in terms of race, gender, class, etc., but at the same time to withdraw yourself from the situation.

Interpretive approach

The interpretive approach incorporates within it the analytic approaches of grounded theory, thematic content analysis, phenomenology, and hermeneutics ([QDA 2011](#), [Mayring 2014](#)). The emphasis is on staying close to the data and interpreting the material from a position of deep empathic understanding. The researcher aims to provide a 'thick description' of the characteristics, processes, transactions and contexts that constitute the phenomenon being studied.

2.   Methodological issues

There are specific issues that need to be considered in developing a qualitative research proposal. A number of the key elements of the traditional research proposal require a different approach in qualitative research. These differences arise out of the specific philosophical background and the nature of the methodology as described above. Issues of sampling, development of research instruments and ethics require particular attention.

Sampling

Given the methods of data collection and analysis, the sample size is usually limited to around 10 to 30 respondents, and is rarely more than 100. Depth and quality of the information provided by respondents are the primary goal, rather than statistical inference. Probability sampling is rarely adopted and then usually to avoid accusation of bias rather than to validate statistical analysis. Small sample sizes require particular assumptions and approaches. Sampling designs are typically based on the specific requirements of your study. There are few direct rules, but the researcher must be transparent as to what decisions were made, and be able to defend those decisions. Convenience sampling, engaging with only those respondents who are easily accessible and cooperative, is strongly discouraged unless there are no practical alternatives. As discussed in chapter 7, even though they will not use the methods of statistical inference, qualitative researchers will often wish to argue the applicability

of their findings to the population from which the sample was drawn and suggest the potential for transferability of those findings to similar populations.

Purposive or strategic sampling

Particular emphasis is placed on *purposive* or *strategic sampling.* Deliberate choices of respondents or settings are made to ensure coverage of the full range of possible characteristics of interest. This should ideally include both those who are seen as typical and those reflecting the diversity of the population being researched. This can be achieved by *stratifying* your sample in terms of:

- Personal characteristics, e.g. gender, age, education, religion, ethnicity;
- Group membership, e.g. occupation, community organisation, political party;
- Social characteristics, e.g. geographical area, class, educational level;
- Experience level, e.g. length or regularity of participation in a given activity.

For example if you were investigating the experience of nurses providing treatment for multi-drug resistant TB (MDRTB), you might look for variation by age, gender, ethnicity, qualifications, training in MDRTB treatment, years of experience in working with MDRTB patients, type of facility, size of population being served by their facility, etc.. Each nurse sampled would include a number of these characteristics. For instance, a young black female staff nurse with five years of experience with MDRTB patients working in a tertiary hospital who has attended an internal two-week course on the treatment regime for MDRTB patients. The aim would be to cover as many combinations as possible, within a given sample size. Usually you would start with respondents that you consider typical of the group. Then you would interview those who are in some way atypical, to see how perspectives differ as you move away from the core group. Other approaches (Patton 1990, Suri 2011) include:

- Extreme sampling - select highly atypical cases (e.g. most/least successful outcomes);
- Intensity sampling - select cases that best illustrate the effects of an intervention;
- Homogeneous samples - select cases in a particular subgroup for in-depth analysis;
- Heterogeneous samples - select diverse cases to identify any common experiences;
- Typical cases - select cases to illustrate the most common outcomes of an intervention;
- Snowball sampling - ask existing informative sample members to recommend others;
- Opportunistic sampling - allow sample design to emerge during fieldwork as suitable respondents are encountered; generally used with hard to reach populations.

Sampling as you go

You can adapt your sample as your research progresses and you learn more about the population, research topic and local context. You need to reflect on your aims and on your stated approach to sampling. It is useful if you are able to analyse your data as you proceed, allowing you to develop hypotheses and/or theories that you can then test. For example, if a common set of views is found amongst a core group you can assess if these same views are held further from the core: do all or only most providers in a facility have positive attitudes to an intervention? Or you can draw in additional people to ask specific questions to test an hypothesis: do junior providers welcome the intervention because they expect to benefit as they gain more experience? Focus on the need to collect a sample that will suit your overall research objectives, while remembering that you will need to justify your decisions and defend yourself against possible accusations of bias - especially if your findings are controversial. There will be a temptation to select respondents that are easier to talk to or who are more friendly and accessible. Where access is very difficult, you must sometimes take who you can get, but you must recognise the risks such an approach poses to the creditability of your findings. Do not change your chosen sample design unless it is either unavoidable or you have a valid scientific reason.

Sample size

There are two theoretical arguments that are widely used to make decisions as to sample size, both concerned with the idea of letting the research process play out. The first, data saturation, involves continuing until no further new information is being uncovered, when additional observations or interviews are tending to produce the same results, even when you look at new categories of respondents (Francis et al. 2010). The second involves forming hypotheses during analysis that become part of later interviews, and continuing until all or enough hypotheses become stable (Thompson 2011).

However, in most cases resource constraints will be the dominant factor in determining sample size. These include cost, time, access and the number, skills and experience of available fieldworkers. Sample size will clearly depend on the length of the interviews. If interviews are an hour or longer then do fewer; if 20-30 minutes then do more. Assuming that interviews are 45-60 minutes, it is seldom necessary to go beyond a sample of around 30 respondents and many studies can be done with 10-15. A sample larger than 60−70 becomes very difficult to manage. Ultimately these are just guidelines. As the researcher you have to decide if your research questions have been answered.

Other methods of data-gathering such as focus groups, participant observation and use of existing documents have their own approaches to sample selection but the key philosophies remain similar.

## 3.	Preparing a discussion schedule

There are a range of research documents that can be used in the research process. These include discussion schedules, observation schedules, workshop agendas and search lists for secondary sources. This section will focus on the development of discussion schedules which can be used in individual interview and focus groups.

The purpose of the discussion schedule is to assist the interviewer to maintain a focus and to ensure coverage of all issues felt to be important. The content of the schedule should be clearly derived from the described research question being addressed and from the stated aims and objectives. Sometimes specialised schedules can be developed for particular purposes such as: directed accounts of behaviour or thought processes; commentary on interventions; or accounts of specific events. You do need to remember that you as the researcher and interviewer are the key research instrument and not the discussion schedule. So while considerable effort needs to be put into the development of the schedule, it should guide the interview and not control it. Thus if you are using other interviewers to collect the data, they need to be thoroughly familiar with the whole study and not just the schedule.

As a guideline, the schedule should not be longer than one or at most two pages, and should be easy to scan quickly during an interview. It is only a guide. The interview does not need to proceed as per the schedule; not everything on the schedule needs to be covered; and the interview can go beyond the items listed in the schedule if the respondent chooses to go into new but relevant areas. Ideally an interview would evolve organically from a single question but in reality it is safer to start with an initial very open question and then to have a limited number of follow-up questions that cover all the components of interest. Each sub-question should be a logical follow-up to the initial question and be accompanied by a checklist of the important topics to be addressed under that heading.

Interviews with specific respondents may need to address additional issues relating to their particular characteristics. If so, these additional points should be noted on the schedule prior to the interview, especially if this person was added to address gaps identified in the existing sample. There are a range of considerations for deciding on the level of detail in a schedule:

- The information being sought;
- Current levels of knowledge on the subject;
- The purpose of the interview;
- The experience of the interviewer;
- The likely openness and ease of the respondent.

These considerations need to be balanced against the need to keep the schedule as brief and clear as possible. Some interviews have particular functions, so each study will require several schedules. For instance, exploring the experiences of providers and patients in a facility. Discussion schedules can be adapted during the research if new questions arise that are seen to be important or new groups emerge that need to be interviewed. However, these do need to be relevant to the original research objectives, or these objectives need to be adapted and the reasons for the changes stated. It is important to maintain this consistency.

4.   Ethical considerations in qualitative research

Three central principles of research ethics are to ensure: respect for participants; justice; and beneficence. Participants will often need or at least feel they need protection from researchers who they perceive to be in a relatively powerful position and able to cause them harm. Particular concerns in qualitative research include:

- Publishing their stories in ways that might put them at risk;
- Undermining their credibility or that of a group that they represent;
- Breaking confidentiality;
- Misrepresenting them;
- Acting against their best interests in some other way.

Access negotiation
Before research fieldwork begins access should be negotiated with all relevant gatekeepers. This is a part of ethical research that is often overlooked or seen simply as a hurdle that needs to be overcome. Working with the community being researched and their representatives is important not only because they are gatekeepers, but for the protection of the research team and because it is simply morally correct. Doing full-access negotiation is particularly important in qualitative research due to the more intensive level of contact and depth of information obtained. Access negotiation is also an ongoing process. While general access is obtained initially, this does need to be repeated at different levels with each new context entered within the site and for each new interview. Clear and honest accounts of the research must be given and an explanation provided of how the data will be used.

Confidentiality
Confidentiality is a right of all respondents. Especially in qualitative research where a lot of detailed information will be obtained from them. Electronic recordings and transcripts of interviews must be password-protected and all personal identifiers removed. Keep the list of respondents separate from the interview data and refer to each using a code number. All recordings should be destroyed at the end of the project, unless storage in an archive has been negotiated. Be careful when writing up the analysis to protect identity.

There are situations where confidentiality may be difficult to maintain. This is primarily where people are interviewed because they hold a particular position, especially if that position is high-profile. Examples include senior officials, doctors or nurses providing specialised services and university professors in particular disciplines. On these occasions it needs to be made clear that while you will try to protect their identity there are risks that some readers of

the final reports or publications will be able to recognise them. Care needs to be taken that respondents are not inadvertently compromised through the research.

Breaking confidentiality

We also have to be aware that in qualitative health research we often collect sensitive information and as such need to act responsibly when confronted by knowledge that indicates that the respondent or others may be at risk of physical or psychological harm. There may be times when we have to break confidentiality. For example when we encounter:

- Respondents who appear depressed to the point of suicide;
- A child or other vulnerable person being abused;
- A respondent who states that they intend to hurt others;
- A respondent who through ignorance or irresponsibility is risking the health of patients, for example by reusing disposable hypodermic needles.

You must warn respondents in advance that you will break confidentiality under these conditions.

Unintentional possible negative impacts

On sensitive topics, for example serious illnesses of children or mental health conditions, you need to give a warning that the interview may evoke an emotional response. Due care and sympathy are required when such a situation arises and if necessary the interview should be ended. Some respondents may ask for help during the interview. If possible, compile a list of agencies that can offer support to those who need it. These agencies should be local and affordable. Consider state services, NGOs and community-based organisations (CBOs) initially. Private agencies may be an appropriate option, but you need to warn the respondent that they will have to pay for these. Bear in mind that the service which is easiest to access may not be the best choice. For example, being referred to a workplace service may lead to others finding out.

In certain situations it may be necessary to arrange for backup services, for example from a qualified doctor or nurse, to be available within your study. You have to be careful to make sure the respondent is not left feeling more vulnerable at the end of the interview, as this may increase the possibility that they may engage in risky behaviour or even be suicidal. You should avoid getting into a counselling relationship with your interviewee and be very careful about giving advice, however tempting, even if it is directly requested. It can increase your power over the respondent and may potentially provide access to information beyond that to which they consented. The respondent may incorrectly assume that you are offering to take some responsibility for their situation and willing to have an ongoing relationship. This is both unethical and bad research. Refer people to those who can provide help.

Consent forms

The fully informed consent of respondents is essential. They must be provided with a detailed explanation of the research and how their information will be used. In addition, specific consent is required if you wish to record or video-record the research interaction. With a child respondent, consent must be obtained from the parents but also from the child. Particular concerns attach to populations who are vulnerable due to their specific characteristics or circumstances, for example the illiterate, prisoners and long-stay hospital patients (Hyder et al. 2004). The respondent is entitled to request at any point that a recorder is turned off and also that a recording be deleted and any notes destroyed. This must be respected. It is completely unacceptable for you to exploit the fact that the respondent may see you as someone in a position to act against their interests if they do not cooperate. If tempted, remember that respondents who feel in any way threatened will typically tell you what they assume you wish to hear rather than what they truly believe.

The process of leaving also has to be negotiated, especially if you feel that you have developed a rapport with the respondent. Deal with any outstanding issues from the interview, such as any painful emotions that may have been released or requests for information that could not be dealt with during the interview. If you wish to correct any misconceptions that arose during the interview this is the time to do it. You should also ensure that promises of follow-up are kept.

Offering incentives for participation

Offering incentives to a potential interviewee may create ethical conflicts. Some argue that it can be considered coercive, depending on the size of the incentive. Others will suggest, especially some respondents, that just taking information without offering anything in return is exploitive. This argument has special force when researchers are seen as obviously more wealthy than their respondents and as profiting from the information gathered, for example in terms of career advancement. The concern often depends on the size of the incentive. One approach is to offer an incentive that will recompense the costs associated with participation, such as the cost of transport to a venue. Incentives for children are often smaller and it is probably advisable to offer gifts such as toys, educational equipment or air time for their cell phones rather than money. One risk is that money given to a child will be appropriated by an accompanying adult, though remember that they may also require compensation if there are costs associated with the child's participation.

Issues for observation research

This may also give rise to ethical concerns as a respondent may unintentionally give away important personal information by their behaviour. You will again have to negotiate access, especially to private venues, usually seeking permission from key gatekeepers. Observational research activities should be in the open, not covert. You can try to blend into the background, but be aware that you may still be seen as an influential outsider - again a position in which you are regarded as potentially having power over others. Those being observed may try to confide in you, for example complaining that they have been waiting for many hours to see a provider. This is not necessarily a problem, but just as in an interview situation you need to be careful how you use information obtained in this way.

5.   Writing a qualitative protocol

While being in some respects a relatively flexible methodology, qualitative research demands considerable rigour if its findings are to be seen as to sufficiently well founded to influence health policy. As discussed in the opening section, a systematic approach is required to ensure the quality of any research activity. This requirement should be reflected in the development of the research protocol, a clear plan of action that should guide the overall research process, including operations in the field, though, as discussed above, it will remain open to adaptation as information is gathered and alternative hypothesis formulated.

Outline and structure
The following sections should be included in the protocol:

- Literature review;
- Research question;
- Aims and Objectives;
- Research design;
- Sample;
- Instruments;
- Analysis;
- Ethics.

Fundamental to a good qualitative proposal is a clear and consistent narrative that flows from the literature review to the description of the analytical procedures that are intended to deliver the findings. The literature review must provide a persuasive introduction to the specific research question to be addressed, with each reference contributing to our understanding of that question. Similarly, the aims and objectives must flow from the statement of the question and the evidence provided by the literature review, though it is possible to define sub-objectives in a study if there are related areas that you want to investigate. The test of the research design is whether it provides, given the sampling procedures and instruments with which it will be implemented, a convincing way of achieving your objectives. Note that if there are questions that you find yourself adding to the instruments that are not in the objectives section, check if these are important and if so go back and amend that section to reflect this.

Literature review

As discussed in detail in chapter 3, this should be a focused account of the literature that covers the subject matter. As indicated above, it should lead to the framing of the research question, addressing relevant international and local contexts and reflecting on theories relating to that question. It will include quantitative studies but the focus will be on the qualitative research literature, and the context within which each study is set needs to be considered. A literature review should be considered the first step in your analysis and so a qualitative analytical approach should be used.

Research question, aims and objectives

The research question, aims and objectives can be seen as a hierarchy, gradually moving from the general to the specific. Within the broad area of work that you are undertaking, for example the overall implementation process for a given intervention, you should first identify the question to which this research activity will contribute, describing the context in which you will be undertaking the research and the relevance of the question to that context. The aims should then clearly specify the areas within the research question that you do plan to address and the objectives the information that you are seeking within each aim. Qualitative research objectives should be formulated in open or descriptive terms, not in terms of measurable variables or hypotheses. Note that specifying clear aims and objectives should not prevent you being open to new ways of thinking about the research question as the study proceed. You can adapt objectives and aims as you gather more information, but it must be a reflective process, with each adaptation involving a revision of the overall protocol to ensure its consistency.

For example, in implementing a new initiative on the provision of antiretroviral drugs, we might be concerned to maximise the take up of the initiative and wish to address a research question such as: 'What are the factors which tend to encourage or discourage greater openness about an individual's HIV status?' We might know that one particularly problematic issue is that of children with HIV. How and when to tell children that they are HIV+ has caused considerable debate. Parents fear the potentially negative impacts on the child, including depression, negative self-worth or even self-destructive behaviour. But if they are not informed there are also concerns about risky behaviour, a lack of motivation to adhere to medication and the possibility that the child will discover their status independently, without the necessary assistance and support. Given that parents are the key decision makers we might define the aims and objectives of a qualitative research project as follows:

Aim:
- To assess how parents make the decision to disclose their HIV+ status to their child.

Objectives:
- To understand parents knowledge and attitudes about HIV;
- To understand how they understand childhood;

- To understand what they feel about their child being HIV+;
- To understand what they understand by disclosure;
- To understand what they see as the risks in telling their child of its HIV status;
- To ascertain how they talk to their child about medication;
- To identify the signals that indicate that they and their child are ready to go through the process of disclosure.

Research design

In relatively small scale qualitative studies, this will often be a relatively brief description of what you want to do, including the basic methodology, and the context within which the research will be undertaken. You may also include more details on some of the specific issues that you wish to look at within each objective. The design section serves as a framework for the technical components that follow. In the above example it might be described as follows:

Individual in-depth interviews will be undertaken with parents who have had different experiences of disclosure or who are preparing to go through a disclosure process with their child. These interviews will be conducted by the researcher who is trained in the methodology and is familiar with core issues around disclosure from reading the literature and from discussions with health staff. All the interviews will be recorded for later transcription. The focus for this study is on parents whose children attend services at a tertiary hospital. The study can later be expanded to include other sites.

Sample

The information required here typically includes the sampling frame, the intended sample size, the sampling procedures to be adopted and details of the types of respondent you are seeking. Note that in observational studies, for example of health facilities, the sample would be specified in terms of the sites that you will observe and the times when you will be observing them. Thus, in the above example we might decide to select 12 sets of parents and to sample them purposively with the assistance of the health staff who provide services to their children. The sample could then be split into those who have already disclosed to their children and those who have not. Subjects might be selected in order to obtain substantial variation on factors such as: age of child; ethnicity of parents; education level of parents; extent of the child's illness; and, for those who have disclosed, the age of the child at the point of disclosure.

Discussion schedule

This section will describe the instruments that you intend to use in your research. It will include a broad description of the specific content that you are looking for and an explanation of the different components of the schedule. A provisional discussion schedule (St John's University undated) or alternative instrument if using another methodology should be attached, with an indication that this will probably be modified as the research proceeds.

Analysis

This should provide a detailed description of the ways in which you plan to analyse the substantial amount of qualitative data that your research will generate (see section 13). It should include a discussion of the preparation of documents, recordings and other materials for analysis, for instance transcription and translation. The process for the identification of themes and the approach to coding, including a description of any software that you intend to use, should be described in some detail. You should also explain how you intend to address the issues of data validity and reliability,

Ethics

This should follow the guidelines described above. The key issues relate to: obtaining the informed consent of respondents, with specific consent for any recordings; obtaining consents for children if they are included among the respondents; confidentiality and the circumstances

under which it might be broken; data protection; any referral plans for respondents who might request or need support; and the rights of respondents to end participation and request destruction of interview notes and recordings at any time.

Ongoing revisions of protocols
You should remember that in qualitative research the protocol is a living document. It can be adapted as you identify new questions or identify answers to some initial questions and want to focus on other issues in later interviews. New target groups can also be identified.

6.   Individual in-depth interviews

Individual interviews (UK Data Service 2016) are the most common form of qualitative data collection. It takes the form of a discussion between an interviewer and interviewee. It involves a conversation but one conducted with intellectual rigour and a clear focus. The interviewer asks the questions, and directs the discussion to make sure the respondent stays on track, but must allow them the freedom to respond in their own terms. You are interested in the interviewee's world as they construct it - not as you construct it. As discussed above, you should be aware that each interview involves a power relationship and you will need to convince the respondent that you really are interested in **learning about their reality and meaning system from their own perspective.** It should allow them space for personal exploration and detailed investigation of their own understandings of the world, encouraged by the interviewer and guided back to the focus of the interview when necessary.

Practical arrangements
In general, the interview should be formally arranged, with a fixed appointment at a suitable time and place, one that the respondent considers safe and appropriate. It can be a great advantage to conduct all of your interviews in a place of your choosing because then you can control the environment and remove any distractions. Of course, you should make sure that all your respondents can easily travel to that location. The interview should start with a clear explanation of the nature and purpose of the research and any material or equipment that will be used. It should preferably be not much longer than an hour in length. You should make sure that you are fully prepared, that you have all your materials with you and that any equipment is fully operational. If using incentives make sure that these are ready. In many cases the availability of suitable refreshments will enhance the mood of the interview and, particularly in poor areas, the provision of food may be a useful way to show empathy with respondents.

Privacy and quiet are very important. The interview must not be interrupted, so telephones should be turned off, those who might come to see you alerted and a 'Do Not Disturb' sign placed on the door. There should be no TV or computer screens, as these are focal points for the eyes. Comfortable upright chairs should be arranged at about a 70 degree angle so you do not look directly at each other and you should avoid having a table or desk in between you. The recorder should be in a location near to both interviewer and interviewee. As indicated above, if possible avoid places that have meaning already loaded into them, for example religious centres, community meeting rooms or doctors' offices. The home of a respondent can be one of the most difficult places to conduct an interview because you are in the position of a guest and have no control over the environment or activities of other household members. Be careful about photographs or wall posters. Even 'innocuous' images such as photographs of political or community leaders, religious messages or health education posters may influence the behaviour of the interviewee.

Recording equipment
Your research may depend on the quality of your recording equipment. There is nothing worse than doing an interview and then not being able to transcribe it. You should seek advice from

someone experienced in the field. Make sure that the microphone is the highest quality that you can afford and that it will record both sides of the interview. Always check equipment before an interview, including the batteries and recording space, and make occasional checks during the interview as quietly and surreptitiously as possible. If the interviewee notices, tell them what you are doing and explain that what they are saying is valuable and you do not want to miss anything that they are saying.

Introduction and preparation
You should respect any concerns expressed by the respondent. If a consent form has been signed, use that to establish clarity. Otherwise, make sure that the respondent has given their verbal consent, that they are still happy to proceed and that they understand that they may stop the interview at any time. If you are recording the interview then make this clear, explaining the right of the interviewee to turn off and delete the recording if they become concerned. This initial phase important to establish trust. Introduce yourself as is appropriate to the interview setting and focus on creating an open and cordial atmosphere and making sure that the interview process is understood. Explain what will happen to the data and how it will be used, emphasising that you will ensure confidentiality as discussed above.

Preparation
As indicated above, the discussion schedule used in the interview should be regarded as a simple checklist of the key points to be addressed. It should not be seen as a rigid agenda that must be followed. Less structure generally allows more openness and free thinking. Given a high degree of flexibility, researchers need to be well prepared not only in terms of the subject matter of the interview and the context, but also in terms of the impression they need to convey in order to encourage the respondent to feel that they can speak freely. If the respondent feels that your questions display little or no understanding of their situation they may feel that there is little point in trying to offer in-depth explanations. If the subject matter is potentially sensitive or shocking then you need to be prepared to control your own emotions and contain those of your respondent, for example if they cry or get angry. Be aware that even a seemingly innocuous question like 'how many children do you have?' can produce an emotional response depending on the situation.

Your personal characteristics, for example gender, age, ethnicity and class, will impact on the nature of the interview, as will the language in which it is conducted, the words used, your dress, non-verbal communications, and emotional displays. You should be yourself, but indicate acceptance of the validity of the respondent's perceptions and attitudes even if you disagree with them. Remember that the interviewee is observing you and using your responses to guide them as in any conversation. You have to continually lead them back into their own reality in order to elicit a successful interview.

When leading a team, it may be helpful to match your interviewers to the respondents in terms of characteristics such as gender, ethnicity, and language, where this is feasible. However, you should ensure that they do not have too much in common. A respondent will probably not want to share personal information with a neighbour or colleague. Access negotiation is very important in these cases to provide a level of trust. It can be very useful for you to meet the respondent initially, so that you can decide which of your team would be most suitable to conduct the interview.

Language and behaviour in the interview
The interviewer should try match dress style, formal or informal, to those of the respondent. Local language usage which is appropriate to the respondent should be adopted (Menaca 2013). Where some term is unclear the interviewer should ask the respondent to explain. If trust has been established they will usually be very willing to help and asking their advice can help balance the power relationship and often results in an increased willingness to share

information. Neutrality and acceptability should be the key watchwords in each interview. You should be yourself, recognising not only commonalities but also differences in perceptions and attitudes. Acting a role very seldom works, especially with those who may be suspicious of your motives. You should be very careful on issues of physical contact, for example shaking hands, as entry into personal space generates its own dynamics and you may well not be aware of the local meanings attached. This also applies to adapting to local conventions on maintaining a physical distance during the interview. Similarly it is useful to remember that your aim is to demonstrate *empathy* (in the sense of indicating that you understand a respondent's position), rather than *sympathy* (in the sense that you endorse that position). There will be times when it is hard to maintain a positive attitude, especially when interviewees are offensive, have made bad choices or test you with aggressive language or actions. You need to be understanding, avoid judgement and know your own limits.

Some writers talk about presenting yourself as naïve to the subject matter but keep in mind, as indicated above, that acting a part may have unexpected consequences. The respondent may guess that you know more than you are saying or alternatively may feel that you have made little effort to understand even the basics of the local situation. You should remember that you *are* always naïve as to the interviewee's reality, even if you are very familiar with the topic under discussion. They are the experts on their own life. You may guess that there is some degree of common understanding between you, but you need to continuously confirm this. One tactic is to treat the interview as generating new perspectives on what you might regard as old realities

Power in the relationship

As in any relationship, power differentials exist in any qualitative research interview. The best results are obtained when power is more evenly distributed. Generally, you as researcher will be perceived by the respondent to have more power. This can generate distrust, so you need to do what you can to counteract this perception. The more power you hand over, the more open the respondent will be but, as in the matter of language and dress, you should be genuine. Respondents will not easily accept a display of pretended servility. Sometimes, power can go the other way. Interviewees may try to dominate, either from common practice or from fear. People such as politicians, officials and senior clinicians are used to playing the dominant role in conversations with their staff and may treat you in the same way. Power can be used to avoid sensitive topics. Especially in the initial interviews you should be aware of your own nervousness. Try to ensure that you do not feel too stressed, for example by avoiding if possible, circumstances which place you at a disadvantage, and know what you are prepared to cope with. You should certainly not be aggressive, but you must be clear that you need to direct the interview when necessary.

7.   Qualitative interview process and techniques

This section explores some of the key tools and processes that you can use or explore during the interview. All can play important roles in developing a successful interview technique and together constitute a strategic approach that you can adapt to meet most contexts. They are presented as simply as possible, but can often prove complex and difficult to implement. Only considerable practice will provide you with the necessary skills.

Interview process

The interview process can be facilitated if your questions follow a natural sequence. Essentially you want to ensure a smooth flow of information that recreate the respondent's world, or this aspect of it, in words. Start with a general open question and follow up with other that cover the broader areas of interest. Avoid whenever possible jumping between questions or from one question to another. Three key techniques for drawing out responses include summarising, clarification questions and the why question.

Summarising: involves reflecting back what the respondent says to provide an immediate check on the interviewer's understanding, and to allow the interviewee to further develop his or her thoughts on the question. You can think of it as reflective listening. This keeps you present in the interview, maintains the flow, and indicates your desire to fully understand the respondent's perceptions and attitudes. It allows you to constantly be doing validity checks to make sure that you are clear about what they are saying. In an ideal interview this would constitute almost all of your contribution.

Clarification questions: serve to get explanations on issues raised by the interviewee about which the researcher is unclear. Especially in new areas of work, unfamiliar terms or acronyms may be used. You need to find a balance between trying to keep the conversation flowing smoothly and your need to know what words mean in order to understand the interviewee. Make clear why you are intervening. For example, in Ghana a number of words are used to describe symptoms that may be associated with what local people describe as *malaria* ([Menaca et al 2013:p5](#)). An interviewer unfamiliar with these terms might say: "I apologise for interrupting your flow of thinking but can you explain how someone who has *malaria* know that she has poa?". Once you have the response, thank the person and then put them back on track with a reflective summary as above.

Why questions: can be used to encourage the respondent to think about an issue at a deeper level. A powerful and incisive tool in the interview process, it pushes the person to look behind their beliefs to what these are based on. For example, "You said that you felt uncomfortable buying drugs for malaria in the local market. Why was that?" However, there is a risk that it will meet with a rebuttal if used inappropriately. Even if you get an interesting response it can cut you off from other information that is more broadly descriptive but important. It should only be used very selectively and not in all interviews.

Use of prompts: There are specific issues that you need to cover in your interview. Ideally most of these would be covered using the systems outlined above, but you may need to raise issues from your discussion schedule that have not arisen, These generally constitute items that you have drawn from the literature or from other interviews as being important and you want to make sure they are discussed even if not spontaneously raised by your respondent. Try to attach these prompts to existing points raised in the interview. When introducing prompts be careful to avoid asking a series of questions as this encourages brief responses.

Onion features of the mind
The onion model is often used to signify the different levels of information and meanings of a person/respondent. On the outer levels are the more superficial levels that are easily shared and are generally shown to the world. Below this are the more personal levels including some of the background and underlying reasons for those superficial presentations. Here some information may even appear to contradict what is expressed more superficially. As you get deeper, more inherent understandings and deeper values are expressed, until at the core is found the basic values that the person uses to define themselves.

Figure 1. The onion model
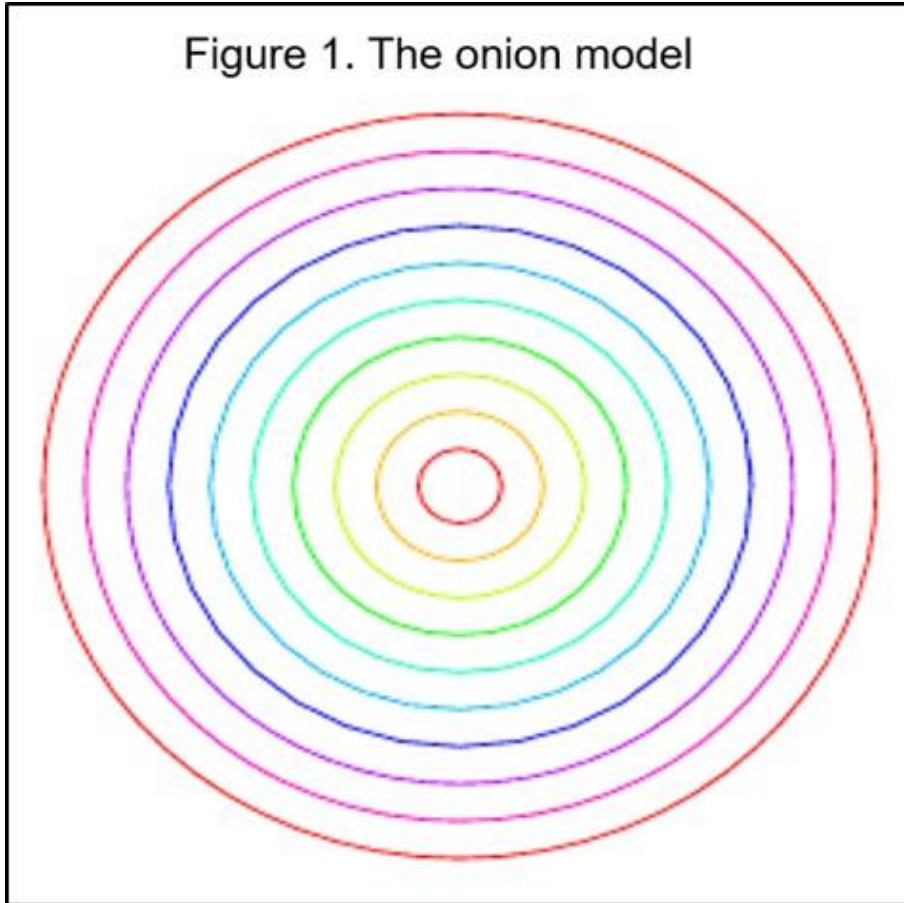
Mind map

A mind map (van Randeraad-van der Zee et al. 2016:1223) is useful as a model, which can evolve as our research proceeds, as to how different items of information, belief, emotion, etc. link together in the mind. So in an interview we may follow one path to get an understanding of how thinking in that area is structured. Then return to another strand of information to explore that. The additional complexity that is more difficult to show is how these links develop interconnections between them. But it still provides a useful model when exploring how a person constructs their information and their narrative on the question(s) that you are posing.

Figure 2: A mind map
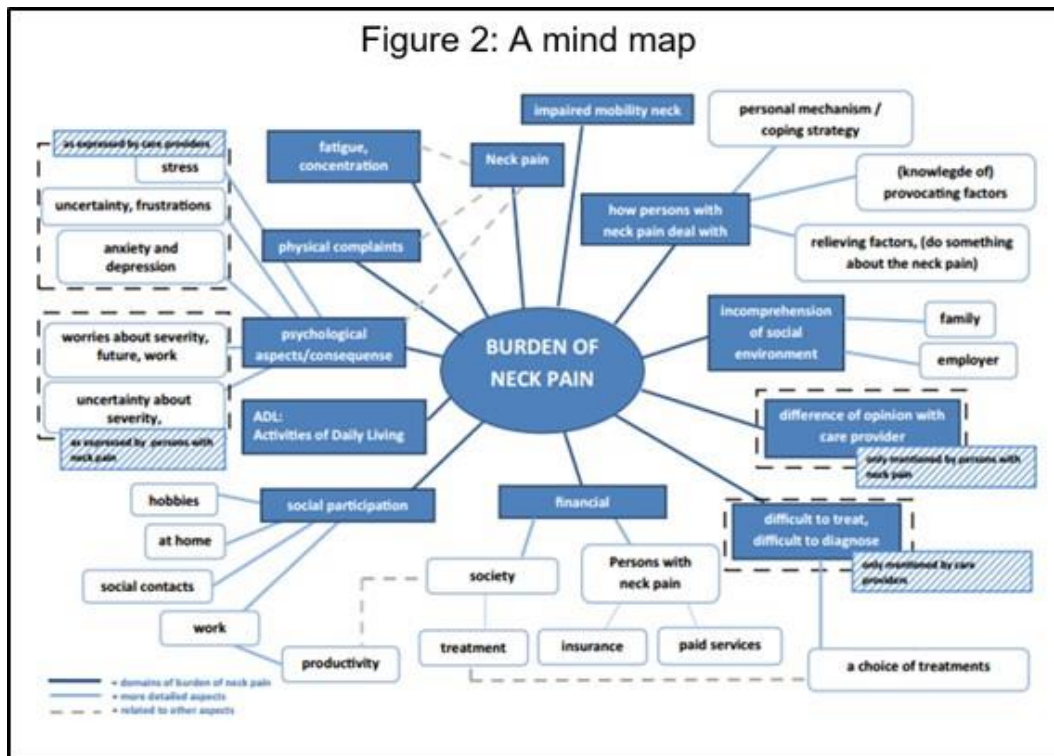
Narratives

The respondent will generally present their information in the form of a story or narrative. It will generally not be a story in which events take place sequentially, as in the accounts we read in autobiographies, but one in which multiple experiences and time periods are intertwined as in most informal conversations. The narrative can be derived from the connecting themes that they use to make it coherent. These may relate to broad constructs such as inequity of treatment, direct experiences such as being disrespected or abused, or hopes and expectations that their situation will improve. Identifying these separate themes within the narrative can greatly enhance the understanding gathered from the interview.

Writing notes during the interview

Writing notes during the interview should usually be kept to a minimum. It moves your attention away from the interviewee, causes breaks in the dialogue, can interrupt the conversation and may cause the interviewee concern. A limited number of short notes can be useful if there is something you really need to remember, for example if a response give rise to an unexpected insight that you fear may not be repeated when you play back the recording. Try to multitask to avoid interruptions and ensure that your notes are organised, legible and will be meaningful when you review them. Have a notebook and pen ready in advance, and explain at the start of the interview to the respondent that you may take notes and why. In particular, stress that these notes are intended to help you in remembering in detail the important issues they are raising and nothing else.

The interviewee's experience

A research interview is often a unique experience for the interviewee; an opportunity to reflect without judgement or outside purpose on an aspect of their lives. To be listened to with rapt attention for an hour and be accepted as having a valid and important point of view no matter what they say. They do not have to reciprocate and there are no clever return comments and if there are incentives you get paid to do it. It can be a novel and pleasing experience. It can also be an undiluted re-experiencing of pain or trauma. If the respondent becomes very upset offer support and maybe take a break. Check when the person is ready to continue and be

sure they are ready and able to carry on. Ending the interview early may not be the best option for the respondent. It may be important for them to tell their story in an accepting environment. As indicated above, remember that you are a researcher, not a counsellor or social worker, and have a referral list of agencies where they can go for assistance.

Some respondents will want to ask questions during the interview. If these are about yourself, be honest but brief, and try to limit any intrusion on the interview. Avoid answering questions about the interview content. Explain that you really want to know what the interviewee thinks. Do not correct a respondent during the interview, even if you feel their perceptions are dangerous to them, such as that HIV does not exist. You can defer answering questions or providing what you regard as correct information until after the interview. Never lie, unless you feel threatened with physical violence. If you do, end the interview in whatever way seems least likely to make the situation worse and leave as quickly as possible.

Ending the interview

You should conclude interviews with care and with respect for the emotions exhibited and ideas raised. Any unresolved personal concerns need to be addressed. As above, refer respondents to outside agencies where necessary. Clarification should be given again as to how the data is to be used, arrangements to ensure confidentiality and access to feedback offered if possible. You can chat informally afterwards, but you should still be careful about self-exposure. Try to minimise talking about any effects that the interview has had on you.

8.   Focus group interviewing

The use of focus groups, as with in-depth interviews and observation, is one more tool in the qualitative toolbox. Each has its specific purposes and applications, and each has its place in the overall research plan. The key value of focus groups relates to their use of group dynamics. Group discussions can stimulate dynamic conversations, which lead to discovery, exploration, direction and in-depth information.

Definition and history

Focus groups are: "*A group of individuals selected and assembled by researchers to discuss and comment on, from personal experience, the topic that is the subject of the research*" (Powell and Single 1996:499). The concept is about 50 years old and, like many modern innovations, its roots date back to the Second World War. A group of sociologists were asked to investigate how the military's propaganda films were being received by their audiences. They learned that, with appropriate encouragement, people could identify the exact reason certain scenes, lines, or phrases made them think or act in a certain way.

The typical focus group consists of a group of five to fifteen participants who gather for a period of 40 to 90 minutes to talk about a prearranged topic under the guidance of a facilitator. It mimics in many respects a group of friends holding a joint conversation while sharing a meal or travelling together, but with the discussion being more focused and with a facilitator mediating the process. The power of the focus group is its capacity to draw out shared knowledge around the key issues being discussed. The discussion format produces information on group or community norms and practices, and may also draw out examples of extreme behaviours and sensationalism (usually associated with people outside the group). The nature of the process may encourage people to share when they may not have done so in another environment (Moosa and Gibbs 2014, Francis and Katz 2013).

Specific roles and use of groups

A focus group produces shared knowledge, with the group building the narrative and information system together. It allows members to confront one another and exchange ideas, building a more coherent story. It is often not the best way to collect sensitive information

(depending on context), unless the group is carefully structured for that purpose. It cannot be used to draw out individual information and stories from all the individuals. One group is not equal to six to ten individual interviews, as is commonly stated.

Focus group structure

There are standard features, but as indicated above they may vary by the demands of the situation. There should be between five and fifteen members, though preferably about seven to twelve. There will be one main facilitator plus a co-facilitator who assists by picking up on issues missed by the facilitator and looking after logistical issues. A session should last about 40-90 minutes depending on the group members. You can break longer sessions in the middle to provide refreshments. The group should be seated in a circle with no other furniture except a small table for the recorder.

Venue

The group should meet by prior arrangement at a fixed time in a suitable location. As described above, the place should be considered non-threatening and appropriate by those attending. Again, privacy and quiet are fundamental. There must be no interruptions, so telephones must be turned off and the location should not be one close to areas frequently used by other people. You should warn those around not to interrupt. There should be no TV or computer screens.   A comfortable upright chair should be provided for each person and the recorder placed near to the (co-)facilitator in the middle of the group.

Recording equipment

High quality equipment is even more important for focus groups as there are many people talking and most will be further from the microphone, which should be one that is designed for this type of purpose. The group size increases the risk of background noise which can make transcription difficult. You should seek advice from someone experienced in the field. As above, always check equipment before the interview, including batteries and recording space, and keep checking at intervals throughout the session.

Selection of members

Selection of focus group members is important. You can select naturally existing groups such as patients attending a clinic on a given day, drug sellers in a local market or members of a community organisation. Alternatively you may first identify a population, for example all those receiving care for a chronic disease at a health facility, and then use information about the members of that population, for example age, gender and ethnicity, as a basis for sampling the number that you need. Selection of members should be strategic and based on the aims and objectives of the study, and on the study context. The key is to be clear as to what questions you feel that the members will be able and willing to usefully address as a group.

There is a difficult balance to be struck between homogeneity and heterogeneity. Market research companies typically use a reasonably large number of focus groups (perhaps 30-40), with individual groups being composed of individuals with much in common but with substantial variation between groups. Homogeneous groups are easier to manage - fewer serious disagreements on controversial issues - and more likely to provide you with a coherent narrative. However, resource constraints will typically strictly limit the number of groups that you can convene and you may need to construct groups that are more heterogeneous in order to reflect the range of perceptions and attitudes that exist in the population. The sensitivity of the content should influence your decisions. For example, in a discussion of risk behaviours in relation to sexually transmitted infections it would be wise to have separate groups for men and women. As indicated above, it is important not to rely on convenience sampling. All selection decisions need to be clearly motivated and recorded.

Before the session

You should develop a clear interview schedule as described in the earlier section, but take note of the particular objectives and structure of the focus group. If you are using others as facilitators make sure they are thoroughly trained in the content and approach. Access negotiation with each of the members of the group is very important and will influence the level of trust. You may want to undertake this task even if you are not facilitating the group yourself. On a practical level, ensure that there are no logistical issues, for example in terms of accessing the venue and travel arrangements for the participants - even apparently minor problems can be disruptive. Prepare as described above for an individual interview: check that consent forms and other necessary documents are available; make sure that you have all necessary equipment with you and that it works; arrange appropriate refreshments; and, if you are using incentives, make sure that these are in place.

At the start of the session

First, go through the consent process in some detail and ensure that the forms are signed. You should have made it clear prior to the meeting that you will be recording the discussion, but it is useful to repeat this before starting. Make sure that the nature of the discussion process is understood and indicate the right of each member to end their participation at any time and have their contribution deleted. Explain what will happen to the data and how it will be used, emphasising that confidentiality will be ensured at each stage.

During the group discussion

Building rapport is your main task as a facilitator. Make sure that you are well prepared in terms of both the topic and the context. An ill-prepared facilitator will not be taken seriously by at least some members of the group. You should try to be relaxed, or at least to appear relaxed, as participants will pick up any anxiety that you display. Try to follow the discussion schedule in your head but refer to it occasionally to make sure that no key issue has been overlooked. Use a general warm-up question to get the group started. Then your main activity is to encourage interaction between the group members, using your background knowledge of the participants or impressions that you have gathered in the informal discussions before starting the session. You should continuously be alert for any pre-existing or developing conflicts in the group if these could hinder interaction. Emphasise that everyone should be allowed the space to make their position clear, even if others strongly object. Your role is as a neutral observer seeking clarity. *Ask, **listen to the responses** and ask.* The facilitator role includes:

- listening;
- observation;
- intervening - but only when necessary to keep the discussion on track;
- encouraging openness;
- displaying sensitivity/empathy;
- supporting/encouraging;
- challenging;
- interpreting;
- modelling - constructing and testing hypotheses as the discussion progresses.

A good facilitator might be described as: curious, has a desire to learn, enjoys asking questions and listens to the answers; outgoing; flexible but persistent; has an open mind; can direct conversations; analytical and sceptical, does not accept answers at face value. While most of us would like to believe that they possess many of these qualities, in fact good facilitators are hard to find. Reflect seriously on the extent to which you fit this description and consider whether the research might benefit from allocating this role to a colleague.

For those acting as facilitators it is important to remember that their personal characteristics will impact on the research process. These will include their gender, ethnicity, social class, language, the words they use, their dress, non-verbal communications and emotions. It will be useful to match the characteristics of interviewers to respondents as far as possible but it can be a serious mistake for them to try to pretend to be someone that they are not. It will be much better if they are honest but indicate openness and respect for others.

Facilitators must be very aware of the power relationship that exists in the focus group. Their task is to hand over leadership to others temporarily not to relinquish it entirely. There are some standard procedures that should be observed. It is important to allow ample time for responses after posing a question. You will probably need to curb dominant participants, draw in quiet ones and act to politely prevent multiple conversations. All contributions should be carefully acknowledged. If pivotal points are raised, take time to make sure that all participants have understood what has been said. Try to avoid *any* expression of your own opinions, including using body language which seems to welcome or reject a contribution. You will need to keep careful track of the process and be aware of both alliances and tensions within the group. Be prepared to break strategically if necessary to avoid disruption. Remember that:

- People sometimes cannot explain why they behave the way they do. Many behaviours are instinctive - not the result of careful consideration.
- Attitudes are complex. They consist of knowledge, perceptions, beliefs, feelings, desires, and opinions buried deep in the subconscious mind. By definition people cannot explain their subconscious.
- Emotions influence behaviour. But again, most people cannot explain their emotions and many prefer to keep sensitive emotions secret.
- Issues of culture are often not directly explainable and the respondents may not even be aware of the principles on which they base their beliefs or behaviours.
- Participants will tend to serve up socially acceptable opinions. They will not want to reveal their inner secrets about sensitive matters particularly in a group format.
- Members are affected by one another and so will respond in terms of the group norms, which may not be true to them.
- Most often participants will talk about the behaviour of others, rather than their own.
- You may get sensationalist responses. These need to be treated with care in the analysis. Often asking for details - who, when, where - may determine the reliability of the response.
- The more sensitive the topic, the more on guard the facilitator should be.

The role of the facilitator can be made much easier if they have an efficient and sensitive co-facilitator. Their role includes: monitoring for people or issues that the facilitator may not be picking up on or giving enough focus to; keeping a check on the group atmosphere, particularly of any changes in mood; writing key ideas down, using the words of participants or paraphrasing for brevity. The co-facilitator should introduce themselves and explain their role at the start of the focus group. While the discussion is taking place they should take responsibility for preventing interruptions from outside and keep a check on the recording equipment. They should debrief with the facilitator immediately after the session.

Hidden helpers may also exist. In most groups there will be a person who will act as an unscripted supporter in maintaining the discussion. If you can identify this person and work with them they can make the process considerably easier. Do not expose them. They will usually not be aware that they are playing this role, but, being generally more sensitive to the issues and group process, will assist in keeping the group on track.

The basic approach mirrors that for individual interviews: general questions, summaries, clarification questions, why questions and prompts, but adapted for the group discussion

context. If the discussion between members is providing the information you need there is no need to intervene. There will almost always be group members who create problems or who need special attention. These include: difficult people who complain and disrupt; those who want to dominate; perpetual commentators; silent members; those who try to ridicule others; and those who passively disrupt the flow of the discussion, for example by making it clear that they cannot wait for the end.

You should end the session with care and with respect for the emotions and ideas raised. Any unresolved emotions need to be addressed. Again, refer individual to outside agencies where necessary. A final clarification should be given as to how the data is to be used, and feedback offered in whatever form you are willing to provide. Note that tt can also often be informative to chat casually with respondents after the focus group in concluded.

Limitations of the method:
Occasionally, focus groups can be of little value, with respondents aiming to please the facilitator or other members of the group rather than offering their own opinions or evaluations. Data is often 'cherry picked' by researchers to support a foregone conclusion. Even commercial market research organisations with substantial resources can have serious failures. The disastrous introduction of 'New Coke' in the 1980s provides a vivid example.

9.   Participant observation

Observation has always been fundamental to health research, for example in the assessment of providers, facilities and treatments (Leonard and Masatu 2005). Qualitative research applications require systematic detailed descriptive observation of behaviours and communications. The researcher can either observe from a distance, or involve themselves directly in the respondents' lives and try to experience at least some part of their reality. Historically this method was one of the earliest forms of qualitative research, linked to ethnography and the anthropological investigations of different cultures. This sometimes involved 'immersion' - living in and interacting on an ongoing basis with a community. The techniques of ethnographic research remain central to the use of observation as a research tool.

Interview-based methods can be seen as more efficient in that the respondents comment directly on their lives and influences. However, the researcher cannot be certain that they are not being misled or simply not understanding what they are told. There are limits to what people will talk about. Observation may provide valuable complementary information, either confirmatory or contradictory. It is based on the assumption that understanding can only be fully achieved by participating in and observing a subject's world. However, the extent to which ultimate immersion take place will involve both strategic and practical decisions by the researcher.

Power of the methodology
By observing over an extended period and becoming part of the life of the observed you may gain access to information that would not otherwise be available. The key issue is that observation is done in their natural setting, where people are much more likely to provide unexpected insights. Those observed in their own reality are more likely to act and speak naturally and in a manner that reflects their true context rather than feeling that they have to respond in ways that are appropriate to a research environment.

Uses
In most examples where observation is used in health research, it will one component of the mix of methods used (Munro et al 2007). Areas of application include the operation of health systems, decision-making processes, disease histories, the spread of epidemics, health risk

behaviours, treatment observation, understanding sites conducive of high risk and, of particular relevance here, the implementation of health interventions. Observation can also be very useful in the context of other data collection activities, for example body language in focus group discussions. Everything that you see could be important and worth reporting in your research diary (the use of research diaries is covered in a later section). Observational data is often ignored as it is perceived as too subjective or too difficult to draw into analysis, but this often leads to vital information being missed. Observation should be core part of every research project as it allows for additional detail to be recorded, even in quantitative research.

Observational data

What is observational data? It is everything that you observe, that you hear and possible that you smell, touch, even taste. It can even include what you do not see - what people are keeping out of sight. It may come in various forms: planned observations of behaviour; observations made during interviews or focus groups; observations of chance interactions; or even secondary sources, where others have described what they have seen. The data usually take the form of written descriptions with some additional sources that may include numbers, e.g. number of patients seen over the period.

10.   Workshops

Workshops are another potentially valuable source of information. The can be defined as facilitated and directed discussion forums. Workshops will typically extend over a longer period than a focus group – sometimes up to several days – so there is more space to explore issues and participants can come and go. They involve a structured discussion process with an agenda and a chairperson. They can include up to 50 or 60 people and still be manageable. Workshops are often set up for purposes other than research, such as planning, discussion of the quality of services or new interventions, community group meetings, negotiations, or education and training. If the subject is of interest to the participants, workshops tend to produce enormous amounts of information. However, to be useful for research purposes, data collection has to be systematic. Workshops will typically form part of a broader research process. The outputs may take multiple forms – formal documents, transcribed notes, behavioural observation. Each of these can contribute to an integrated analysis. It is a methodology that has been used extensively by NGOs and CBOs.

Specific practical arrangements are required. Workshops are very different from the focus group in that respondents often need to come prepared on the issues to be discussed and be working towards specific goals. The venue needs to be large enough for all the participants to come together for some parts of the discussion, preferably with breakaway rooms for smaller group meetings. A strong chair(s) and scribes are required to facilitate discussion and take notes. A clear agenda is needed, preferably sent out in advance. It is often useful to get participants to prepare and possibly circulate inputs before the meeting. Catering is often required, especially if the meeting continues over a full day or more.

Selection of participants

The nature of the workshop will often define the participants. Usually it is people and or representatives of organisations that are affected by, involved in or knowledgeable about the issue under discussion. Selection is purposive with the aim of putting together the most appropriate combination of participants. Usually only one or two workshops will be held so there is a focus on getting the right people. Care must be taken about the mix so that serious conflicts do not occur or one group inhibit another.

Facilitation and recording data

Facilitating these meetings requires considerable skills in addition to those indicated above, mainly learned from experience. At the start you should explain your proposed role in the

meeting, even if you have previously discussed this when the participants were invited, and the use you intend to make of the final documents, with special reference to arrangements to ensure confidentiality. You should respond to any questions that participants may wish to raise. Given the length of time that a workshop takes, and the comparative size, it is not always practical to record and transcribe a full workshop. This may depend on the relative importance of the event. The most common data are the workshop minutes, jointly created documents, copies of the prepared inputs from speakers and notes from any small groups. The scribes should also take more detailed notes of the discussions, with another possibly asked to take notes on group dynamics. Audio recordings may be made but will typically cover only a small part of the proceedings, given the need for later transcription. Discussions with participants during breaks may also provide information that has not been raised publicly. The validity of the outputs should be confirmed by constantly checking back with selected participants.

Workshop series

The workshop series builds on this approach to implement a staged process, allowing time for reflection between each stage. The first workshop takes discussion to a point where some level of clarity has emerged. This discussion is summarised by facilitators, distributed prior to the next workshop and used as a basis for further discussion. Additional investigations may be undertaken in the interim to provide further information. This process can continue until a full resolution is obtained, or a specified number of meetings has been held. For example, a situation review of a clinic in which there is high patient dissatisfaction may go through several cycles to identify all the problems. In a policy review it may be helpful to focus sequentially on different issues such as content, resources and implementation. Additional information or the need for additional participants to adequately address an issue may be identified in an earlier workshop and then included later. Sometimes the complexity of the issues at hand may require periods of extended contemplation between workshops.

Delphi technique

An adaptation of the Delphi technique could be one type of workshop series. This possibility would arise in the situation where the researcher has to find out 'the truth' or at least some shared account between two or more groups of informants. Sessions are repeated with the divergent groups, separately or together, where conflicting ideas are confronted and some shared position found, or some evidence sought. This process continues until there is agreement or at least sufficient shared clarity.

Workshop series can be conceived in terms of a social action model, which implies a process of community development or social change. In its simplest form, a problem is identified and the community is brought together to find a solution. Ideas are tested and then at a later fixed time, the community meets again to decide the next steps. This is repeated until a solution is found. This methodology has proved very successful as a way of drawing divergent stakeholder groups together in a constructive debate.

11.   Additional methods of data collection

Qualitative research methodology is flexible and open to adaptation. You can use your imagination to develop innovative ways to collect relevant information and understand the reality of your target populations. However, you should operate within the overall philosophy and conform to ethical practice. These are some examples that are found in the literature or that the author has explored or developed, but they do not represent the full universe of options.

Case studies

This is more an approach involving the compilation of data on a small number of 'entities' that are studied in depth, rather than a specific method. It has a long history in health research.

There are some famous case studies of individual patients such as those reported at length by Sigmund Freud. But the approach can also be applied to facilities, groups of patients and communities. It usually involves the integration of data from multiple sources, including interviews, focus groups, observation and existing documentation that may have been designed for other purposes, combined with quantitative data, for example on financial resources, costs and service utilisation. It can be a very cost-effective way to study initial implementation processes for interventions, particularly when the selected cases are located in very different contexts.

> Example: A prevention of mother to child transmission (PMTCT) programme is to be implemented. The implementation research will focus on levels of utilisation and the quality of services and try to identify barriers to access and bottlenecks in the implementation process. Case studies of a small number of purposively selected clinics might involve: observation of prospective mothers at meetings informing them of the new service and in the clinics during return visits; interviews with mothers, nurses and counsellors; review of statistics and documentation on adherence to the regime for antenatal care; following a cohort of women over time from testing through delivery to completion of treatment.

This approach can quickly expose problems such as: areas of risk, e.g. not entering programme due to lack of understanding; areas of potential breakdown, e.g. queues for service and attitude of staff; lack of clarity about interventions in contexts, e.g. informed consent to test; contextual issues, e.g. stigma; and additional concerns, e.g. knowledge and lifestyle factors. It is especially useful for complex interventions or contexts where it is difficult to take all potential influences into account. It is an often criticised approach, primarily as findings cannot be readily generalised, but that is not the only role of research. Description is important and can be highly enlightening. This is where the case study approach is strongest. It can provide the space to explore new ideas and is describe processes.

Case studies can be made more rigorous. You can undertake a longitudinal study, rather than relying on recall, use standardised measures and find ways to check the validity and reliability of the data used, perhaps using an independent analyst. Check for data sources that may have been missed, especially those that might show a different outcome. Finally, be clear on the limitations of the research and document these. Closeness to the data can create a false sense of certainty. Get an independent colleague to critically review your findings.

Role play and visualisation exercises

Role play involves creating a situation in the research space where subjects are able to play out a real life situation (Alberta Health Services 2016). It has been widely used in health education initiatives but also has the potential to provide valuable research insights. Visualisation exercises (IDS undated) create a similar effect, but focus on the responses to visual images (Kennedy et al 2016). With young children observation of play and interpretation of drawings can also elicit useful information. This approach requires a high level of skill and knowledge, especially in terms of 'debriefing' the subjects once the exercise is complete to understand their perceptions and behaviours. These approaches are useful in situations where the information you seek may be suppressed, for example when patients feel uneasy about criticising providers, or in situations where the data are sensitive and the person will find talking uncomfortable (Bogart et al. 2013). These are intensive exercises that require considerable time. They also necessitate careful access negotiations, particularly around the notion of informed consent, as you are seeking information that the person may be hiding, intentionally or unintentionally.

Expert review

Implementation research will very often require detailed policy analysis and a review of technical documentation, such as treatment protocols. It is a mistake to assume that these

activities can be undertaken by researchers with limited expertise either in the methodology or the subject matter. Submitting the documents to a small group of experts, with terms of reference stating exactly how they are meant to comment, for an agreed fee, may seem extravagant but be a cost-effective alternative. Select your experts carefully, according to their proven expertise and trustworthiness. The sample of experts is usually small, four to ten people, as they will be expensive. Analysis of their reports involves a modified content analysis, as discussed below in section 13, with you providing continuity across their various findings.

Narrative interviews

Narrative interviews have a long history in health research. They have been used to understand institutional development, experiences of illness and treatment (Groleau et al. 2006), and community histories (Walker et al. 2014). Subjects are asked to recount their life story as a whole, or to talk about particular periods, for example of illness. It is usually done through interviews, but often more than one - sometimes up to five or more. Stories are often backed up by pictures, newspaper stories, other published accounts, friends' stories, or objects of importance from a particular time. The District Six Museum is a very interesting example of such collections of information. Confidentiality will vary according to the purpose of the interview. Multiple people who share a common experience will often be interviewed, either separately or together, to find a shared understanding of events (Lucas et al. 2009). The analysis aims to clarify a period of time or event in the context of this person or multiple people's lives.

Specialist techniques

Sometimes where it is difficult to access stories or memories, innovative techniques can be useful. There are several online participatory method toolkits that may provide inspiration (IDS undated), FAO undated).

Body scans/maps (Gascaldo et al. 2012): have been used with trauma survivors (Riaño-Alcalá et al. 2010:39) and in studies of reproductive health (Cornwall 1992). An outline of the subject's body is traced on a large sheet of paper. They then write on the paper linking different parts of the body with pains that they experience, good or bad associations, related past and future events, and why they like or dislike that part of themselves. Putting it on paper provides a distancing, sometimes allowing the person to talk about their life and experiences with less inhibition.

Rope and stones: is one of a range of approaches used to reconstruct life stories (Whalley 2016). It was very successfully used with refugee children who had often travelled for more than a year (Ruf et al. 2010). Even before migration their lives had been unsettled and their memories of traumatic events were confused between what they had experienced in their home country, on their travels and in their current location. A rope was laid on the ground to represents the child's life. Stones were used to represent bad events, sticks other important changes and buttons positive events. Arranging these objects was often a shared task, with the child leading and other children and family members or friends assisting, facilitated by the researcher. The method allows the child and the researcher to agree on the order of events. The child can then tell their story.

Memory boxes: were developed both as a therapeutic approach for grief and a way of gathering information around the HIV epidemic, for families where parents with young children were dying (Denis et al. 2003). A memory box would store the parent's life story and messages from the parent to the child. It would usually include a book with a written story, photographs, print media where available, favourite objects, books, etc. It was primarily used as a family tool for passing on memories of parents, but has been used as a research tool to really understand the lives of those families affected by AIDS. At a simpler level, scrapbooks could form a similar function in a different context.

Photographic essays or posters: involve respondents being given a camera and then asked to go out and take photographs that hold meaning for them around a particular theme ([Photovoice undated](#)). The photographs are organised into a narrative or used to create a poster around the theme, for example HIV prevention, health services, nutrition, etc. The researcher will interview the respondents using these outputs to understand their perceptions and attitudes.

12. Final thoughts on data collection

You should be creative in your approach. Qualitative methods are flexible and can be adapted as necessary to obtain required information. Remember the key rules of inquiry: respect your participants and stay within ethical rules and guidelines; be aware of the impact of different subjectivities on the data being gathered; and be aware that some methodologies require additional levels of skill and/or new technologies and equipment. Be ambitious but honest with yourself about gaps in your expertise and experience.

Keeping journals of observation and of your own experience
The power of observation has already been noted but it does need to be systematic observation. Maintaining a formal research journal can facilitate this and should be used to document the overall research process, being updated each evening. Everything that you see and hear could be significant, but for the journal you need to make decisions as to what could be most important for the project. You should also record your own reflections and those of any colleagues involved in the research. Observation is sometimes perceived as too subjective or too difficult to draw into analysis, but this often leads to vital information being missed. Remember that the researcher is the research instrument, documenting the world as it is seen. Subjectivity and awareness of what is shown and not shown are key.

Research diaries
Three diaries should be kept, especially by less experienced researchers. These are less formal than the research journal and each has a different focus and role.

The first diary is of observations in the field. This literally means recording anything you see of interest, including the physical context, people that you see, events that happen, notes from conversations or even the impact of the other senses. It is not only what you see, but if you are working with fieldworkers or know others familiar with the context, recording their insights. It is mainly an aid to memory. The entries will consist of short notes, each with a date and time of day.

The second diary will contain your ongoing analysis and the development of ideas, interpretations and hypotheses. It will usually be written up during periods of inactivity and provides a chronological record as to how your ideas developed. It is for your use only and therefore can include crossings out, side notes and personal reflections.

Especially for less experienced researchers it is also useful to keep a personal diary. It will be used: to record lessons learned; the impact of interviews or experiences, emotionally, intellectually, spiritually or even physically; ideas for self-improvement; technical notes for future research or interviews; and questions or concerns that need further reflection. It is useful for your development as a research instrument and to promote the valuable habit of reflexivity.

13.   Preparing for analysis

The key to analysis of qualitative materials is to work slowly and carefully. If you do not allow enough time or are not relaxed, you risk making serious errors of judgement. Remember that these materials are all you have to make sense of your experiences in the field. You need to give them your full attention.

Transcribing and preparing for formal analysis
Taped material should be transcribed (Dresing et al. 2015) and where necessary translated. The task of transcribing should not be underestimated. It will take approximately six hours of transcribing for one hour of recording, nine hours if translation is required. This is where you start to value the extra money you spent getting a good-quality recorder. Use this transcription process to reflect on your own interviewing technique and skill. Pages of transcription need to be prepared for analysis. They should be done in a standard text format, not in a table. Use 1½ line spacing and have a wide margin, at least 7cm on the right.

Decide on the level of detail for the transcription. You need to make sure at a minimum that all words are included. Decide whether to add: notes on body language; interruptions; time taken in pauses; speed of speech; and other expressed emotions, such as crying, laughing or shows of irritation or anger. Make sure that it reads easily and grammatically, if translated. If transcribed in the original language, do not correct the grammar.

Add a brief introduction to each transcript to provide background. This can assist in contextualising the interview during analysis. Include: personal characteristics of respondent; some background and reasons for their selection if purposive; the context of the interview, including place and time; notes on other events during the interview, e.g. interruptions, external noises; any specific notes on the interview, e.g. overall attitude, body language, signs of anxiety; and notes on yourself while undertaking the interview, e.g. tired, emotionally affected, angry, excited. Both the voice of the interviewer and that of the respondent need to be transcribed. Validity checks should be carried out on the transcription. Best practice is for the person who did the interview to do the transcription and for a colleague to validate the transcription using the summaries described in the interview process.

Where you have to translate as well as transcribe, it is generally better to do both at the same time. Doing both together allows for a meaningful translation rather than a direct translation. You will need to take into account that words and even more so sentences can have alternative meanings in every language. If you conducted the interview using a translator, everything from all three participants has to be translated and transcribed. This is important, as translators working in the field will often miss some of the details of both your questions and the subject's responses.

Computer assistance
You can purchase specialist equipment but this is not essential. Transcribing programmes can be downloaded from the web at no cost and work reasonably well. Express Scribe is one programme that the author has used successfully. The media player on MicroSoft Office can also be used. Some researchers prefer to purchase a foot pedal control to advance and stop the recording.

14.   Analysis and interpretation of qualitative data

Qualitative analysis involves extracting meanings from the texts and other materials that you have compiled to generate persuasive narratives (Pope et al 2000). A range of analytical techniques have been developed (Taylor and Gibbs 2010). This section will outline an

approach adopted by the author, which can be described as *contextualised interpretive content analysis*. Note that data analysis does not require to be undertaken as a strict sequence of distinct steps. Different aspects of the analysis will overlap and you can constantly backtrack on earlier decisions, and even review your original aims and objectives. However, the final version should reflect a systematic and coherent process of evidence-based analysis and interpretation.

Process of data analysis

In qualitative research analysis should start from the conceptualisation of the project, with an increasing focus over time, as the data are assimilated, on synthesis and conclusions. Many important insights will come in the early phases of research and these should not be lost. You should keep reviewing your research journal and diaries. In an open-ended project you can always go back into the field if you are not satisfied with your findings. The key requirements if for 'immersion' - becoming totally familiar with your data. If you have done your own interviews and transcription you should already know the content well. If not it is even more important to read and reread the transcripts so that you can get beyond the linear account. Read your field notes and diaries field alongside the interview transcripts, such that they become a cohesive unit.

Identifying themes

From your knowledge of the material you can identify themes, "recurrent unifying concepts or statements about the subject of inquiry" (Bradley et al 2007:p1761), preferably arising out of the data, but also those that you will want to impose based on your research question (Atkinson and Abu El Hap 1996). The former approach will provide a rich description of the totality of your data, while the latter allows you to focus in on selected aspects. These themes are the building blocks for extracting meaning. Use the language of the interviewees to name the themes. Imposed themes will usually be derived from existing theories, but you can still represent them using the words of the research subjects. You should try to go beyond simple descriptive terms to look at processes such as the relationships between concepts and the power dynamics that underlie perceptions, attitudes or behaviours (Vaismoradi et al. 2016). Keep your aims and objectives in focus. The themes need to reflect back to these but remember that you can still adjust aims and objectives to take new findings into account.

It is worth spending time on this exercise as it will guide the rest of your analysis. Test your initial system of themes by reviewing a selection of interviews. Keep the word length of themes short, at most a few words, as this will make the coding and analysis easier. Add definitions to each of the themes to ensure accurate coding. Remember that this is just an initial phase of the analytic process. You still have to interpret the data, but this will determine the constructs underlying the analysis. The next step is to organise your themes. Ideally you need to end up with about 40 to 60 themes, classified under some five to eight major themes. For example, in a recent study on community knowledge and attitudes regarding non-malaria febrile illnesses in Eastern Tanzania, Chipwaza et al. (2014) categorised their data using the following major themes:

1. Participants' understanding of fever.
2. Awareness of the community on non-malaria febrile illnesses.
3. Treatment seeking behaviours for febrile illnesses.
4. Health workers' practices towards non malaria febrile illnesses.
5. Capacity for diagnosis and management of febrile illnesses in health facilities.

Under the first major theme, they identified the themes:

1. No knowledge of exact meaning
2. Illness: malaria, colic, rheumatism, sleeping sickness
3. Symptoms: 'hot body', headache, coughing, rashes, body pain

4. Causes related to illness - measles, tuberculosis (TB), typhoid fever
5. Other causes - change in weather, sunlight, etc.

During the process of identifying themes, you should also be developing 'research memos'. These are notes that you make to yourself, just as you have been doing in your research diaries. They may be on points of interest, possible alternative meanings of a section of text in the transcript, ideas for further analyses, or hypotheses that you may want to test at a later stage in the analysis. Memos can be attached to individual pieces of text or be general notes to yourself.

Elaborating

Elaborating is the process of making connections across different sections of a single transcript or across multiple transcripts. As an interview progresses key themes may be returned to several times. Connections need to be made between these points as each return can offer a new perspective. Connections across transcripts also allow for different perspectives, but you must also take into account the different contexts. This occurs once you have immersed yourself in your data and are looking at quotes across interviews out of the sequence in which they originally appeared. It provides an opportunity for reflection and, where necessary, expanding your analysis. There is a delicate balance to be struck as the researcher needs to distance themselves from the text, while being aware of their own subjective relationships to the content

Coding

Eventually you should be able to establish a reasonably stable set of themes, which can be used to code the data (DeCuir-Gunby 2011). This involves re-reading all of the transcripts and attaching themes to sections of text. This may be a line, sentence, paragraph or several paragraphs. It can and often will be necessary to assign multiple codes, as any piece of text can carry multiple meanings. The sections may overlap, so you can code a paragraph with one theme and then code one sentence inside it with another.

As you are coding, new themes or areas of interest may arise that you need to add to your list, or new meanings may need to be attached to existing themes. This may necessitate you going back to recode earlier transcripts. There are different systems of coding:

- Computer-based coding (provided in the analysis software discussed below);
- Writing in the margins (similar but less control, and can become untidy and confusing);
- Colour coding (but this requires a large number of distinct colours and it is difficult to identify overlapping themes);
- Cutting and pasting (again it is difficult to deal with overlapping themes and you risk misinterpreting text that is taken out of context).

Working with a computer program is therefore strongly recommended. Software makes coding simple and easy to modify, overlaps are clear and it allows for a systematic cut and paste function (in this case the interview name is listed with each quote so the context is maintained). Two of the most widely used programs are Atlas.ti and Nvivo. These can provide both systematic control over your text, themes and memos and the flexibility that will allow you to change and develop themes and coding systems as you go along. They are also very useful in terms of validity checking, as you can review all the quotes on any selected theme. However, you still have to read and define your themes, undertake the coding and draw out the analysis. They are reasonably expensive, and like all software programs, are constantly being developed and improved.

Interpretation and establishing conclusions

Now all the pieces are in place for you to undertake an in-depth interpretation. Your hypotheses, inductions that you have developed as you went through the thematic content analysis, need to be tested against the data and accepted, modified or rejected. If your theme development was done well, and reflected your original aims, then these themes will become the focus for interpretation, though perhaps not exclusively, and the section titles of your findings will reflect this.

Interpretation has to happen at multiple levels. To illustrate, each of your headed sections should provide a discussion as to how this area of the content was addressed and interpretations made. You should try to present the reader with a coherent narrative account of the relevant research material. You should usually start with a discussion as to how the main theme is understood, and then look at variations and elaborations around this. Quotes should be used to illustrate key points. In the conclusions these points have to be drawn together into a more coherent analysis. Again, it is essential to establish a distance between the researcher and the text in order to be able to reflect on the content. There should also be an acknowledgement of limitations and the possible impact of these.

Incorporating context

It is generally agreed that a detailed knowledge of context is important in understanding what people say. Context works on multiple levels. In reading the transcript of an interview, statements need to be considered in terms of what else the person is saying at that time and in relation to the general subject matter of the interview. For example, parents may respond quite differently to questions relating to traditional medicine depending on whether the interview relates to the health of their child or themselves. The location in which an interview was undertaken, for instance in a clinic or their home, should be taken into account when interpreting the responses of a patient, as should their personal characteristics, for example their gender, age, race and educational background. It may also be important to consider the emotional state of the respondent. Questions about the quality of health services may receive very different responses if the subject has recent painful experience of the death of a loved one in hospital. Finally, remember that the interviewer is also part of the context and that their characteristics, such as race and gender, level of skill, state of mind, etc., must also be considered.

Verifying interpretation and testing validity

Qualitative methods has produced new challenges in terms of demonstrating validity. Some tools such as inter-rater reliability, with the data being independently coded and the codings compared, are overvalued and outdated (Armstrong et al. 1997). Such approaches do not work well with qualitative methods due the freedom of thinking and interpretation that is encouraged. A more useful approach is to set up theories and then search the data for contradictions. When these occur, you can use that information to modify the theories and then continue looking at more data until these stabilise.

Having an audit trail is important so that the process of reaching conclusions is fully documented. Triangulation, for example using the findings from a reanalysis of existing quantitative data, may provide a useful guide. Validity can also be enhanced by drawing in additional people, either by getting other researchers to look at the same materials, or by getting the community from which the respondents were drawn to comment on the findings. Such approaches do however have to take into account the role of subjectivity. External validity can be explored by considering if the theory developed or the interpretations made have a value in explaining aspects of the wider world. You can also assess how others, for example fellow scientists, officials and health providers working in the field of study respond to the results. The assessment of such stakeholders as to the quality of the description of the

context and research process will often be the key to looking at the potential for applying the knowledge gained to other settings.

As emphasised above, it is essential to recognise your own influence, from the setting of the research question to the final conclusions. This is true no matter what the methodology but the impact can be considerably greater in qualitative research, given that it embraces subjectivity. Acknowledgement of the influence of the researcher can range from a detailed discussion to a brief statement. It is good practice to state upfront the position from which you approached the research and then in the conclusions to again state what influences you may have had on the findings.

Remember you are not doing inferior quantitative research. Numbers are really only useful when describing your research methods, for example to indicate the number of interviews or focus groups conducted and the numbers of different types of respondent in your sample. Even statements like 'most felt this' and 'a smaller number felt that' should be avoided. The role of the researcher is to provide insights - not descriptive summaries. The other crutch adopted by some researchers is to fix on a particular analytical approach and apply it slavishly. This will give you a narrow perspective. The analysis process should be creative and flexible, and must be treated as such.

At the same time, make sure that the interpretations that you make are genuinely grounded in the data. You can make conjectures, but you must identify these as such and stress that they will need further research. You should also avoid individualising a text, by getting too involved with a few respondents and allowing your findings to be dominated by the perceptions, attitudes and behaviours of those individuals. Similarly, you should not get personally involved with the text, for example by attacking or expressing your admiration for an individual. You can strongly disagree with the position they have taken, for example advocating the use of herbal remedies to treat AIDS, while avoiding personal criticism.

Research report

The stated aims of the research should be the focus of the final report. This is where the ordered flow of your analysis really facilitates the presentation of findings. Each major theme can act as a section heading. The sub-themes under each become important discussion points. The conclusions draw these together to make a final statement. Finding the right balance between discussion and quotes is sometimes hard. Especially when the material is interesting there is a tendency to add more quotes and for the findings to be dominated by these. There is also a false belief that by just using quotes bias is removed. The researcher has the responsibility to lead the analysis and present the findings. Quotes are there to illustrate these conclusions and to contribute to the narrative. As a rule of thumb there should be at least twice as much discussion text as there are quotes.

References

Albarracin D, Johnson BT, Fishbein M, Muellerleile PA (2001) Theories of reasoned action and planned behavior as models of condom use: A meta-analysis. Psychological Bulletin 127: 142-161. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4780418/

Alberta Health Services (2016). Teaching Sexual Health. http://teachers.teachingsexualhealth.ca/teaching-tools/instructional-methods/role-play/

Atkins, Salla, Willem Odendaal, Simon Lewin, Henry Lucas and Mariano Salazan (2015), Qualitative Evaluation in Health Care – online course. ARCADE. http://courses.arcade-project.org/course/view.php?id=2

Armstrong, David, Ann Gosling, Josh Weinman and Theresa Martaeu (1997). The place of inter-rater reliability in qualitative research: an empirical study. Sociology 31(3):597-606. www.academia.edu/458025/The_place_of_inter-rater_reliability_in_qualitative_research_an_empirical_study

Atkinson, Sarah and Monica Abu El Hap (1996). How to do (or not to do)...Domain analysis for qualitative public health data. Health Policy and Planning 11(4): 438-442. https://heapol.oxfordjournals.org/content/11/4/438.full.pdf

Bogart, Laura M, Donald Skinner, Idia B Thurston, Yoesrie Toefy, David J Klein, Caroline H Hu and Mark A. Schuster (2013). Let's Talk! A South African Worksite-Based HIV Prevention Parenting Program. Journal of Adolescent Health 53. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3707983/

Bradley, Elizabeth H., Leslie A. Curry, and Kelly J. Devers (2007). Qualitative Data Analysis for Health Services Research. Health Services Research 42(4):1758-1772. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1955280/pdf/hesr0042-1758.pdf

Chambers, Lori A., Sergio Rueda, D. Nico Baker, Michael G. Wilson, Rachel Deutsch, Elmira Raeifar, Sean B. Rourke (2015). Stigma, HIV and health: a qualitative synthesis. BMC Public Health 15:848. http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-2197-0

Chang, Sun Ju, Suyoung Choi, Se-An Kim and Misoon Song (2014). Intervention Strategies Based on Information-Motivation-Behavioral Skills Model for Health Behavior Change: A Systematic Review. Asian Nursing Research 8(3):172-181. www.sciencedirect.com/science/article/pii/S1976131714000504

Chipwaza, Beatrice, Joseph P. Mugasa, Iddy Mayumana, Mbaraka Amuri, Christina Makungu, and Paul S. Gwakisa (2014). Community Knowledge and Attitudes and Health Workers' Practices regarding Non-malaria Febrile Illnesses in Eastern Tanzania. PLOS Neglected Tropical Diseases 8(5):e2896. http://journals.plos.org/plosntds/article/asset?id=10.1371%2Fjournal.pntd.0002896.PDF

Cooper S and Endacott R. Generic qualitative research: a design for qualitative research in emergency care? Emergency Medicine Journal 24(12):816-819. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658349/pdf/816.pdf

Cornwall, Andrea (1992). Body mapping in health RRA/PRA. RRA Notes 16:69-76, IIED London. http://pubs.iied.org/pdfs/G01449.pdf

Daniels, Karen, Rene Loewenson, Asha George, Natasha Howard, Gergana Koleva, Simon Lewin, Simon et al. (2016). Fair publication of qualitative research in health systems: a call by health policy and systems researchers. International Journal for Equity in Health 15(1):1-9. http://equityhealthj.biomedcentral.com/articles/10.1186/s12939-016-0368-y

DeCuir-Gunby, Jessica T, Patricia L. Marshall and Allison W. McCulloch (2011). Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project. Field Methods 23(2):136-155. https://www.researchgate.net/publication/254091978_Developing_and_Using_a_Codebook_for_the_Analysis_of_Interview_Data_An_Example_from_a_Professional_Development_Research_Project

Denis, Philippe, Sibongile Mafu and Nokhaya Makiwane. Memory box programme: training manual. Handicap International. http://www.asksource.info/resources/memory-box-programme-training-manual

Denzin, N.K. and Lincoln, Y.S. (eds) (2000). Handbook of Qualitative Research 2nd edition. Sage. http://nersp.nerdc.ufl.edu/~ufruss/documents/ryanandbernard.pdf

Dresing, Thorsten/Thorsten Pehl, Christian Schmieder (2015). Manual on Transcription. Transcription Conventions, Software Guides and Practical Hints for Qualitative Researchers. 3rd English edition. Marburg 2015. http://www.audiotranskription.de/english/transcription-practicalguide.htm

FAO (undated). PRA Toolbox. Conducting a PRA Training and Modifying PRA Tools to Your Needs FAO. http://www.fao.org/docrep/003/x5996e/x5996e06.htm

Fossey, Ellie, Carol Harvey, Fiona McDermott, Larry Davidson (2002). Understanding and evaluating qualitative research. Australian and New Zealand Journal of Psychiatry 36:717-732. http://pathways.bangor.ac.uk/fossey-et-al-evaluating-qual-research.pdf

Francis, Shelley A and Mira L Katz (2013). The HPV Vaccine: A Comparison of Focus Groups Conducted in South Africa and Ohio Appalachia. Maternal and Child Health Journal 17(7): 1222-1229. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3799762/pdf/nihms521364.pdf

Francis, Jill J, Marie Johnston, Clare Robertson, Liz Glidewell, Vikki Entwistle, Martin P. Eccles and Jeremy M. Grimshaw (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychology & Health 25(10): 1229-1245. http://aura.abdn.ac.uk/bitstream/2164/2654/1/Francis%202009.pdf

Freire, P. (1970). Pedagogy of the Oppressed. New York: Seabury. http://www.msu.ac.zw/elearning/material/1335344125freire_pedagogy_of_the_oppresed.pdf

Gale Nicola K, Gemma Heath, Elaine Cameron, Sabina Rashid and Sabi Redwood (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Medical Research Methodology 13:117. http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-117

Taylor, C and Gibbs, G R (2010). What is Qualitative Data Analysis (QDA)?
http://onlineqda.hud.ac.uk/Intro_QDA/what_is_qda.php

Groleau D, Young A and Kirmayer LJ (2006). The McGill Illness Narrative Interview (MINI): An Interview Schedule to Elicit Meanings and Modes of Reasoning Related to Illness Experience. *Transcultural Psychiatry* 43(4):671-691.
http://ltc-ead.nutes.ufrj.br/constructore/objetos/MINI.pdf

Howarth, Caroline and Foster, Juliet and Dorrer, Nike (2004) Exploring the potential of the theory of social representations in community-based health research - and vice versa? *Journal of health psychology* 9 (2): 229-245.
https://core.ac.uk/download/files/67/93122.pdf

Hyder, A A, S A Wali, A N Khan, N B Teoh, N E Kass, L Dawson (2004). Ethical review of health research: a perspective from developing country researchers. *Journal of Medical Ethics* 30:68-72.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1757137/pdf/v030p00068.pdf

IDS (undated). Participatory Methods. Institute of Development Studies.
http://www.participatorymethods.org/resources/type/tool-12

Kennedy, Anne, Ivaylo Vassilev, Elizabeth James and Anne Rogers (2016). Implementing a social network intervention designed to enhance and diversify support for people with long-term conditions. A qualitative study, *Implementation Science* 11(27):1-15.
http://implementationscience.biomedcentral.com/articles/10.1186/s13012-016-0384-8

Lambert, Carol, Julie Jomeen and Wilfred McSherry (2010). *British Journal of Midwifery* 18(5):321-326.
http://www2.hull.ac.uk/student/PDF/GraduateSchoolBJM_18_5_reflexivity.pdf

Leonard, L Kenneth and Melkiory C. Masatu (2005). The use of direct clinician observation and vignettes for health services quality evaluation in developing countries. *Social Science and Medicine* 61(9):1944-51.
http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1248797108129/Leonard_TheUseofDirectClinician.pdf

Lucas, Henry Shijun Ding and Gerald Bloom (2009). What do we mean by 'major illness'? The need for new approaches to research on the impact of ill-health on poverty, in *Health and social protection: experiences from Cambodia, China and Lao PDR*. Bruno Meessen, Xiaomei Pei, Bart Criel and Gerald Bloom, eds. ITGPress, Antwerp.
http://www.eldis.org/go/home&id=39284&type=Document#.V4jzA7grIdV

Lukobo MD and Bailey RC (2007). Acceptability of male circumcision for prevention of HIV infection in Zambia. *AIDS Care* 19(4):471-7.
http://www.ncbi.nlm.nih.gov/pubmed/17453585

Mack, Natasha, Cynthia Woodsong, Kathleen M. MacQueen, Greg Guest, Emily Namey (2005). Qualitative Research Methods: A Data Collector's Field Guide. *Family Health International*.
https://www.fhi360.org/sites/default/files/media/documents/Qualitative%20Research%20Methods%20-%20A%20Data%20Collector's%20Field%20Guide.pdf

May, Maria (2012). RCTs: Not All That Glitters Is Gold. *Stanford Social Innovation Review* 28 August.

Mayring, Philipp (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt. 2014. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173

Menaca, Aranta, Christopher Pell, Lucinda Manda-Taylor, Samuel Chatio, Nana A Afrah, Florence Were et al. (2013). Local illness concepts and their relevance for the prevention and control of malaria during pregnancy in Ghana, Kenya and Malawi: findings from a comparative qualitative study. *Malaria Journal* 12:257. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3724599/

Moosa, S & A Gibbs (2014). A focus group study on primary health care in Johannesburg Health District: "We are just pushing numbers". *South African Family Practice* 56(2): 147-152. http://www.tandfonline.com/doi/pdf/10.1080/20786204.2014.10855353

Munro, Salla A, Simon A Lewin, Helen J Smith, Mark E. Engel, Atle Fretheim, Jimmy Volmink (2007). Patient Adherence to Tuberculosis Treatment: A Systematic Review of Qualitative Research. *PLoS Medicine* 4(7):1230-1245. http://journals.plos.org/plosmedicine/article/asset?id=10.1371/journal.pmed.0040238.PDF

NIH (undated). E-Source: Behavioral and Social Sciences Research. http://www.esourceresearch.org/portals/0/uploads/documents/public/glanz_fullchapter.pdf

Nnko S, Boerma JT, Urassa M, Mwaluko G, Zaba B (2004). Secretive females or swaggering males? An assessment of the quality of sexual partnership reporting in rural Tanzania. *Social Science and Medicine* 59(2):299-310. http://researchonline.lshtm.ac.uk/14751/

Patton, M. (1990). *Qualitative evaluation and research methods* (pp. 169-186). Beverly Hills, CA: Sage. http://legacy.oise.utoronto.ca/research/field-centres/ross/ctl1014/Patton1990.pdf

PHAST (2011). Health Knowledge. Department of Health UK. http://www.healthknowledge.org.uk/public-health-textbook/medical-sociology-policy-econo

Photovoice (undated). Reframing the world. https://photovoice.org/matt-daw-participation-in-development/

Pope, Catherine, Sue Ziebland, Nicholas Mays (2000). Qualitative research in health care: Analysing qualitative data. *BMJ* 320:114-6. http://www.brown.uk.com/teaching/HEST5001/pope.pdf

Powell, R.A. and Single, H.M. (1996) 'Focus Groups', *International Journal of Quality in Health Care* 8(5): 499−504. http://intqhc.oxfordjournals.org/content/intqhc/8/5/499.full.pdf

van Randeraad-van der Zee, Carlijn H., Anna J. H. M. Beurskens, Raymond A. H. M. Swinkels, Jan J. M. Pool, Roy W. Batterham, Richard H. Osborne, Henrica C. W. de Vet (2016). The burden of neck pain: its meaning for persons with neck pain and healthcare providers, explored by concept mapping. *Quality of Life Research* 25(5):1219-1225. http://link.springer.com/article/10.1007/s11136-015-1149-6

Riaño-Alcalá. Pikar (2010). *Remembering and narrating conflict*. Historical Memory Commission of the Center of Memory of Colombia and the University of British

Columbia. http://www.reconstructinghistoricalmemory.com/wp-content/uploads/2013/05/RememberingNarratingConflict_120313V5.pdf

Ruf, Martina, Maggie Schauer, Frank Neuner, Claudia Catani, Elisabeth Schauer, Thomas Elbert (2010). Narrative Exposure Therapy for 7- to 16-year-olds: A Randomized Controlled Trial With Traumatized Refugee Children. *Journal of Traumatic Stress* (23)4: 437-445. www.researchgate.net/publication/45495356_Narrative_Exposure_Therapy_for_7-to_16-year-olds_A_Randomized_Controlled_Trial_With_Traumatized_Refugee_Children

Schwartz-Shea, Peregrin and Dvora Yanow (2012). *Interpretive Research Design: Concepts and Processes.* Routledge Series on Interpretive Methods. Routledge. http://samples.sainsburysebooks.co.uk/9781136993831_sample_827038.pdf

Sibbald, Bonnie and Martin Roland (1998). Understanding controlled trials: Why are randomised controlled trials important? *BMJ* 316:201. http://www.bmj.com/content/316/7126/201

Sikkema, K.J., Watt, M.H., Meade, C.S., Ranby, K.W., Kalichman, S.C., Skinner, D. and Pieterse, D. (2011). Mental Health and HIV Sexual Risk Behaviour Among Patrons of Alcohol Serving Venues in Cape Town, South Africa. *Journal of Acquired Immune Deficiency Syndromes* 57(3):230−37. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3135683/

Skinner, D. (2000). An Evaluation of a Set of TRC Public Hearings in Worcester: A small Rural Community in South Africa. *Psychology, Health and Medicine* 5(1):97−106. https://www.researchgate.net/publication/233487204_An_evaluation_of_a_set_of_TRC_public_hearings_in_Worcester_A_small_rural_community_in_South_Africa

Skinner, D., Tsheko, N., Mtero-Munyati, S., Segwabe, M., Chibatamoto, P., Mfecane, S.; Chandiwana, B., Nkomo, N. and Tlou, S. (2006). Towards a Definition of Orphaned and Vulnerable Children. *AIDS and Behaviour* 10(6):619-27. https://www.researchgate.net/publication/7140611_Towards_a_Definition_of_Orphaned_and_Vulnerable_Children

Serekoane, Motsaathebe, Carla Sharp, Donald Skinner, Lochner Marais (2014). Untold Stories of Fieldworkers Working Amid Adverse Conditions. *Journal of Ethnographic & Qualitative Research* 8(3):157-172. https://www.researchgate.net/publication/262805109_Untold_stories_of_fieldworkers_working_amid_adverse_conditions

St John's University (undated). Developing an interview schedule. http://www.sjut.ac.tz/policies&Forms/Developing%20an%20Interview%20Schedule.pdf

Strecher, Victor J, Brenda McEvoy DeVellis, Marshall H. Becker, Irwin M. Rosenstock (1986). The Role of Self-Efficacy in Achieving Health Behavior Change. *Health Education Quarterly* 13(1):73-91. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/68171/10.1177?sequence=2

Suri, Harsh (2011). Purposeful Sampling in Qualitative Research Synthesis. *Qualitative Research Journal* 11(2)63-75. http://dro.deakin.edu.au/eserv/DU:30064369/suri-purposefulsampling-postprint-2011.pdf

Thomson, S. B. (2011). Sample Size and Grounded Theory. *JOAAG* 5(1). http://www.joaag.com/uploads/5_1__Research_Note_1_Thomson.pdf

UK Data Service (2016). Teaching Resources: Interviewing Methods. ESRC. https://www.ukdataservice.ac.uk/teaching-resources/interview

Vaismoradi, Mojtaba, Jacqueline Jones, Hannele Turunen, Sherrill Snelgrove (2016). Theme development in qualitative content analysis and thematic analysis. Journal of Nursing Education and Practice 6(5):100-110. https://www.researchgate.net/publication/290552438_Theme_development_in_qualitative_content_analysis_and_thematic_analysis

Walker, Melissa, Fredericks, Bronwyn, Mills, Kyly, & Anderson, Debra (2013). "Yarning" as a method for community-based health research with indigenous women: the indigenous women's wellness research program. Health Care for Women International 35(1): 1216-1226. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.660.4814&rep=rep1&type=pdf

Watt, Melissa H., Stephen M. Kimani, Donald Skinner, Christina S. Meade (2016). "Nothing Is Free": A Qualitative Study of Sex Trading Among Methamphetamine Users in Cape Town, South Africa. *Archives of Sexual Behavior* 45(4):923-933 http://s3.amazonaws.com/academia.edu.documents/44856768/Nothing_Is_Free_A_Qualitative_Study_of_S20160418-11761-ei32ql.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1467041704&Signature=PaRrlUXD%2FBlUFTlbaMcbGvGY4w8%3D&response-content-disposition=inline%3B filename%3DNothing_Is_Free_A_Qualitative_Study_of.pdf

Watt, M.H., Aunon, F.M., Skinner, D., Sikkema, K.J., Kalichman, S.C. and Pieterse, D. (2012). "Because He has Bought for Her, He Wants to Sleep with Her": Alcohol as a Currency for Sexual Exchange in South African Drinking Venues. *Social Science and Medicine* 74(7):1005-12. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3298605/

Chapter 10: Participatory research

Pauline Oosterhoff and Danny Burns
Institute of Development Studies

1.   Introduction

Participatory methodology in health systems always involves members of a local community. However the term 'community' can be used in many different ways. It can applied to administrative or geographical units or social, even global, categories. People who suffer from a medical disorder are often seen as a community by those who do not share their condition, including doctors, but may not see themselves as a community. People with HIV, diabetes or another chronic disease may share little besides that disease. Different definitions and interpretations of the term community has led to diverse expectations, perceptions and experiences amongst different stakeholders, and confusion as to the findings from program implementation, evaluation and research (Mansuri and Rao 2003, Schmittdiel 2010).

The social, administrative and local boundaries of communities therefore need to be carefully defined at the outset of a health intervention and used consistently during research in order to avoid misunderstanding and facilitate measurement and interpretation. What a health system is - and by extension what it is not - also needs to be established. As discussed in chapter 1, for some it includes a variety of informal health sector actors including unqualified providers, religious healers, patient organizations, and mobile pharmacies on the back of a bicycle. Others limit it to state sanctioned formal actors. This chapter is not about discussing the merits of alternative definitions but it is important to specify these before embarking on participatory research, as they are essential in delineating the context and identifying stakeholders.

2.   What is Participatory Research?

In this section we discuss some of features that distinguish participatory research. First and foremost, it requires the active engagement of participants in the key decisions about the research trajectory and research design as well as in all of the analytical components. Participatory approaches are relevant for implementation and scaling up not only in terms of generating data, but also because participatory methods generate internal learning by different stakeholders - including beneficiaries - which can enhance political and social sustainability. The following are key characteristics of participatory research:

The ***research process should benefit those engaging in the research process*** as well as others seeking to gain knowledge from the research. This can complement and enrich well established ethical principles in research of ***beneficence*** and ***autonomy***: research must make a positive contribution towards the welfare of people and respect and protect the rights and dignity of participants.

Participation should extend as far as possible throughout the research cycle: determination of research questions; research methods; research design; data collection; data analysis; proposed solutions and action. Some so-called participatory research is little more than a way of getting local people to do data collection for researchers. When the participatory label is used in situations where local people feel that they have to engage in processes over which they have little if any control it can be 'a new form of tyranny' (Cooke and Kothari 2001) Sometimes questions may reasonably be framed by outsiders, but in this case they need to be genuinely owned by participants to be seen as participatory. Similarly, participatory facilitators may know from past experience what research methods might be most likely to succeed, but participants have to understand those methods, and want to adopt them. These

requirements imply that participatory research in health systems will typically necessitate a long term active engagement by researchers with in-depth knowledge of the local situation.

Participatory research recognizes multiple perspectives. The stakeholders engaged in participatory research should represent different (interest) groups in a health system, including those in need of services from the site in the country where the implementation takes place. Who should be involved depends on the questions that are being asked. Mapping techniques (Barker 2013) are often used to determine the key stakeholders and the power relations between them. Participatory research is based on discerning the appropriateness of interventions through triangulation of the multiple subjectivities of these diverse stakeholders. It is critical of universal claims about reality. Rather, it is based on an assumption that medical and social realities may differ and that effective health interventions also need to take social realities into account. Health experts, for example, might use positivist methods to test the efficacy of a vaccine, while more participatory methods are used to assess the appropriateness of the intervention to the population and the efficacy of implementation.

It has long been established by medical anthropologists that health care problems such as patient dissatisfaction, inequity of access to care, and spiralling costs are often not amenable to traditional administrative or biomedical solutions (Kleinman et al. 1978). Failure to take local realities into account often results in poor outcomes from investments intended to improve population health. Long distances to newly installed water points can put young girls and women at risk of sexual violence (UNHCR 2015). Lack of perceived ownership may result in the dismantling of expensive new sanitation blocks for parts. In many instances, the best technical solution is not necessarily the best social solution (Burns and Worsley 2015).

Ethical participatory research recognizes power and diversity. Rather than assuming 'universal' western models and philosophies of health care and health systems, participatory research acknowledges power relations and the implications of the historical diversity of practices within countries. This encourages critical reflection and supports meaningful implementation of ethical research principles such as informed consent. A focus on individual rights and consent in health system research, for example, can be meaningless in cultural contexts where strategic life and health-related choices and decisions are made in a familial and/or community context (Oosterhoff et al 2008; Oosterhoff 2009). When people feel they are under economic pressure, it can cause them to participate for financial reasons (Lindegger et al. 2006). Without such a proper understanding of local contexts, research may serve mainly to reinforce institutional and structural power inequalities rather than operate in the interests of either subject or researcher.

Similarly, ***participants in participatory research do not regard themselves as separate from the subject of the research***. As Loewenson et al. (2014) point out, Participatory Action Research:

> *"transforms the role of those usually participating as subjects of research and involves them instead as active researchers and agents of change. Those affected by the problem are the primary source of information and the primary actors in generating, validating and using the knowledge for action"* (Loewenson et al. 2014:12)

3.   Is participatory research always qualitative?

Sometimes people position participatory research as a form of qualitative research that is in opposition to quantitative research. This is not the case. The defining issue for participatory research is who decides what data to collect and who analyses and interprets that data - not what sort of data it is. What follows from this is that while there are a number of specific methods that have been particularly associated with participatory research, a wide variety of

data collection methods, including sample surveys, focus groups, key informant interviews and ethnographic studies, and a similar range of analytical approaches, for example statistical modelling and thematic analysis, can be integrated into a participatory research process. Research methods that are commonly associated with participatory methods include:

Peer research

Where researchers are trained to support a group of individuals like themselves to generate data, undertake analysis and produce outputs. This is particularly useful for vulnerable groups who are unlikely to 'open up' in the presence of 'authority figures'.

> Example: In 2003 Burns and colleagues supported a facilitated research process with young people in Hounslow (Percy-Smith et al. 2003), an area of relative deprivation in London. They worked with the local community health council to identify people in the area from a range of very different settings and backgrounds. They trained this core research group in participatory inquiry approaches. Each of the groups then went back to their communities and explored what the key health issues were. The groups produced a range of outputs using various creative methods including poems and story boards, as well as more traditional analytical techniques. The groups were brought together into a learning workshop alongside grass roots practitioners (including teachers, midwives and health visitors) and senior managers (for example, Directors of Education and Health services). They were able through visual methods (see below) to collectively identify priorities that could be fed into the development of local health policy. This research "challenged health professionals to reflect on their own assumptions and practices in response to the complex health concerns of young people and the need to connect more effectively with the real lives of young people in research and policy development. The collaborative action enquiry approach used models an 'alternative' and arguably more effective approach to policy learning involving young people, providing further evidence of the value of action research in health sector research" (Percy-Smith 2007:891)

Action Research

There is a long history of Action Research which ranges from reflective practise with an individual focus (which has a strong history in nursing practice - for examples see Vallenga et al. 2009) to co-operative inquiry with a group focus; to Participatory Action Research (Baum et al. 2006) (which tends to have a community focus - for examples see EQUINET: TARSC and ALAMES 2014) and Systemic Action Research (Burns 2014a, Burns 2014b, Burns 2007). Action Research traditionally involves a process where groups explore issues in a cyclical way: starting by assessing the situation; then planning; then taking action; then assessing the effect of that action; in the light of this reassessing the situation; and so on. Action research in health systems can adopt both participatory and non-participatory research methods. In North East India a group of international and national public health researchers looked at the opportunities for indigenous Khasi women to participate in health policy making in the area of Sexual and Reproductive Health and Rights. The research methods included both non-participatory methods such as desk research, individual and group interviews by local indigenous researchers, and digital story telling (Oosterhoff et al. 2015a). Key in the approach was that the results were shared with many stakeholders, both policy makers and local Khasi women's groups, as part of a joint analysis.

PRA (Participatory Rural Appraisal) methods

Rapid appraisals are well established public health tools in a wide range of development and emergency settings (UNHCR, WHO 2008; Needle et al. 2003; Annett and Rifkin 1995). In particular, they will often be used in contexts requiring Rapid Appraisal and Response (RAR), a terminology used to highlight the need for timely effective action in response to the findings of such an appraisal. Note that it is quite common for rapid appraisals to employ local or marginalized groups such as young people, drug users or urban slum dwellers. That does not

necessarily mean they are participatory - unless these people are also involved in the analysis of the results. Genuine participatory research approaches must enable local people to identify their own priorities and make their own decisions about actions that should be taken based on the results. A range of methods for rapid appraisal were developed by Robert Chambers (1981) and others in the 1980's. They can include tools such as transect walks, participant observation, mapping, preference rankings, Venn diagramming (chapatti diagrams), village mapping, body mapping, life stories, counting and classifying essential and commercial medicines within households, and approaches that work with mystery patients or consumers (Mohanan et al. 2015).

Visual and performative participatory methods

These include participatory video, digital stories, photo-voice and participatory theatre for development. Visual methods are a good way of identifying what really matters to people, triggering emotional responses, opening up dialogic inquiry into issues, and communicating effectively what the issues 'look like'. For example, a photo-voice exercise on the prevention of HIV and AIDS in South Africa uses photographs of a crowded taxi to show where the dangers of transmission lie in day-to-day life. Similarly, they show pictures of buildings without windows which make visceral the realities of the health environment (SLF, 2015). Theatre for Development in Nigeria is:

> *"increasingly used to explore health problems and their causes, for example in reproductive health. It recognises the inherently conflictual nature of interests, relations, and power around some of these issues. The dramas not only serve as codes for collectively identifying health risks and determinants, they also contribute to the shift in power relations needed to address these determinants or to support demand for or uptake of services. The process of building the drama often simulates collective conflict, to facilitate reflection by those who have power as well as to empower those who lack it"* (Cornwall and Jewkes 1995).

Participatory mapping

This includes: systemic causal mapping - which might be used to show the causal dynamics of how disease is spread; GIS neighbourhood mapping - which can be used simply to show the location of health facilities and catchment areas, or as a dynamic digital tool identifying hotspots in patterns of disease or accidents (Burns and Worsley 2015, Tatern et al. 2012, Emch et al. 2012), and social network mapping - which might be used to support discussions about how behaviour change might spread (Igras et al. 2016). Social mapping can also be used as an underpinning for participatory statistics, both to collect data and as a way to display or organize an analysis (Oosterhoff et al. 2016).

Participatory statistics

Many participatory processes will generate numbers, for example rankings and counts, and there will often be a joint analysis - involving multiple stakeholders - of the numbers collected. However, they typically make limited use of traditional statistical methods. Researchers adopting an approach which has come to be called 'Participatory Statistics' (Chambers 2007, Holland 2015) may in addition use standard statistical principles such as probability sampling frames and statistical power calculations to sample and code data which can be non-participatory but is rooted in a participatory process. For example, local people may collect health data at village level, which they analyse and interpret before it is aggregated, following standard statistical procedures, for analysis at higher levels (Riemenschneider et al. 2013).

4. How might participatory research be relevant when changes in the delivery or financing of health services are being proposed?

Participatory research methods allow technical knowledge to be integrated with local social knowledge in a deliberative process which enables effective intervention strategies to be developed. They can be used in many disciplines by various professional groups that study health systems: medical doctors, nurses, patient organizations, medical anthropologists and other social scientists.

There is a wealth of research, for example, on the low uptake of sexual and reproductive health services due to a lack of understanding of social and cultural realities. One instance relates to an international project aiming to improve maternal health among the Black Thai, an ethnic minority group in Northern Vietnam, which failed to recruit 'traditional birth attendants' because they assumed these would be women. Among Black Thai it is actually the men who - together with close family or neighbours - traditionally assist women with child birth (Oosterhoff et al. 2011). In North East India, national health policies made in the capital led health providers to provide contraceptives to indigenous Khasi which were not only medically dubious - if not harmful - but socially unacceptable (Oosterhoff et al. 2015b).

When changes to health services are introduced, participatory methods can help to understand where these should best be made in order that people will actually use them. For example, to plan for HIV and AIDS related referral systems in Vietnam, health managers, practitioners and people living with HIV (PLHIV) were asked to draw up maps to show what services people with HIV used. The maps of the three groups were totally different. Health policy managers assumed for example that people would use free testing services close to home. PLHIV however said that anonymity and speed of test results were their main criteria. The services they actually went to were literally not on the map of the health service providers.

Participatory research can also help to assess if any change should be made. Vietnam had an opt-out HIV testing service when global policies, including that of WHO recommended offering anonymous and opt-in testing. Opt-out was seen as compulsory and possibly a violation of human rights. Action research actually showed that opt-in testing was not preferred because it would require patients to ask for it and health staff to offer it to some - but not all - patients. Both health staff and patients wanted to avoid the stigma and shame attached to the disease, which was seen as less of a problem under an opt-out service (Oosterhoff et al. 2008).

Participatory approaches can also help management to understand and detect differences in service provision, and set priorities. The authors have used participatory approaches in health policy work (as discussed in the Hounslow example above) and public health leadership work. For example a year-long action research programme with Directors of Public Health explored the decisions and dilemmas faced by senior managers in health. This group learning process bought to the surface a huge diversity of approaches and priorities of public health leaders and enabled them to learn from each other. (Burns et al. 2004a; Burns et al. 2004b).

Loewenson et al. (2014) cite a wide range of examples of the uses of Participatory Action Research in Health Systems. For example: Generating risk maps in a steel mill in Mexico; building a community-based child health information system in South Africa; analysing narratives from community outreach workers in Chicago; developing a participatory poverty index in China; participatory 'pharmocovigilance' to identify and evaluate previously unreported adverse drug reactions in Uganda; and community research to support primary health care in Zimbabwe. The extent of the examples, and the co-publication of this book by WHO, indicates a widespread recognition of the value of participatory studies alongside more mainstream health research.

What participatory research methods might be most appropriate?

There is a wealth of information available on participatory methods in different contexts. As it is an approach that can be used in many disciplines, and there are dozens of ways of classifying and categorizing methods and tools, it is worthwhile to have a fairly clear idea about the question that one wants to address before starting an internet search. A lot of tools are just about collecting data with beneficiaries who may participate in feedback meetings or consultations, but will not be involved in the analysis. A useful place to start is the participatory methods website, which provides resources to generate ideas and action for inclusive development and social change, including in the field of healthcare. A participatory approach to communities infected and affected by HIV and to PLHIV and their support groups has been key in HIV and AIDS research. See, for example, UNAIDS and Royal Tropical Institute (2004). If you would like to know more about action research and participatory networks you may also want to look at the ALARA network.

5.   The use and value of participatory research

While participatory research has had a long tradition of generating rich data for local 'meaning making', the process by which individuals interpret their situation, relationships, events, etc., it has struggled against the critique that its analysis is only of value in a specific local context and cannot be used to draw more generalised conclusions. Over the past 10 years or so methods have advanced considerably (Burns and Worsley 2015). Systemic Action Research (Burns 2014a, Burns 2014b, Burns 2007) enables multiple inquiry streams to operate in parallel across a wide geographical terrain; collective analysis processes (Easpaig 2015) allow large amounts of narrative data to be analysed by local communities within a matter of days rather than weeks; participatory statistics enables verification of emerging hypotheses in ways which mirror those used in the analysis of traditional sample surveys.

In the example that follows we show how the use of a participatory methodology within a multi-method research programme on slavery and bonded labour opened up profound questions about local health systems (Oosterhoff et al. 2016). Experience derived from the research programme can be used to illustrate: (a) how participatory methods can be used at scale; (b) how mixed participatory methods generating both qualitative and quantitative data can be used to support and triangulate each other; and (c) how significant health issues can be revealed through such participatory processes.

Slavery and Bonded Labour in Northern India: A case study of mixed methods participatory research

Since 2014, the authors have been co-directing a research programme to provide data and analysis in support of NGOs and others working to combat slavery. Modern slavery can take many forms and there are many definitions. That adopted by the Freedom Fund, a philanthropic initiative dedicated to the fight against modern slavery and a partner in the programme, is that that individuals in slavery: are paid nothing or below subsistence wages; cannot walk away; and are subjected to violence or threats. More generally, they are used, controlled and exploited by another person for commercial and personal gain.

The slavery work is in three intervention sites. Two are in India: one in Uttar Pradesh and Bihar, focusing on brick kilns, stone quarries and sex work; and one in Tamil Nadu, focusing on cotton mills. The third is in the South Western Terai, Nepal, and focuses on agricultural bonded labour. Although the same research design has been used in each, here we discuss the Northern India research. This has three building blocks:

(1) Life story collection and collective analysis using the techniques of participatory clustering and causal mapping

(2) Participatory statistics
(3) Action Research

These three are interconnected. Over a period of approximately three months, each NGO collected some 44 stories from the villages where they worked, resulting in 355 stories. In each village they collected seven stories of people who were assessed as being in slavery, two from people who had been at risk of slavery but avoided it, and two from leaders or professionals in the village who had some insight into the slavery situation. Prior to story collection, grass roots NGO workers were trained how to encourage people to tell their stories in ways which would not tend to bias those narratives, and how to ask follow up questions which deepened the issues which the story tellers had raised.

The collective analysis workshop allowed analysis of 353 life stories over 4 days. First, participants were divided into pairs. Each pair had approximately 20 stories to analyse. They were given simple questions. What is the primary message that this story is telling us? What are the most important factors in the story and why? A participatory clustering exercise was then undertaken. The major themes identified for each story were shown on a board and participants co-related similar themes. The findings from this exercise were that there were over 50 stories which focused on loans for health and a further 20 which focused on other health related issues. Participants were then instructed in the basic concepts of causal mapping and asked to construct a 'causal map', to plot what they saw as the causal links between the various factors associated with slavery, using the evidence from the stories they had analysed. This was a collective activity involving all participants and the map, covering some 20 sheets of flip chart paper, was displayed on a wall in order that everyone could follow its construction and contribute.

Figure 1: The causal mapping process in action
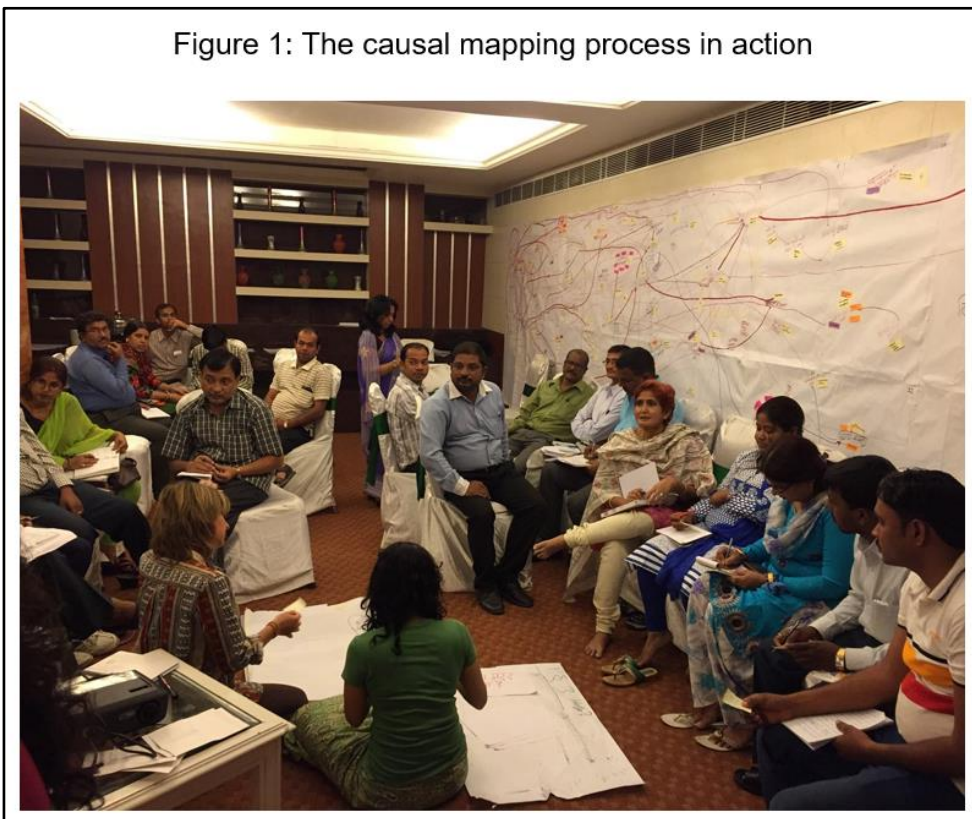
The causal map revealed a pattern which was essentially as follows. People who live in bondage have little or no money because they receive below subsistence incomes. When they hit a health crisis they need money to pay for care. Because the banks will not give money to

slaves, and the local credit and loan schemes cannot provide enough, they have to go to a middle man who will broker a loan (often from a land owner) at interest rates of 60-120% per year. Because they have no money, they guarantee their own or their child's labour for two to three years to pay back the loan - either directly to the money lender or to an associate. The work is often physically hard and hazardous labour carried out with little protection, leaving workers prone to accidents. Poor food and housing conditions also increase the risk of disease. Both may lead the family to need additional money for health care, deepening and prolonging indebtedness.

Figure 2: A simplified causal map
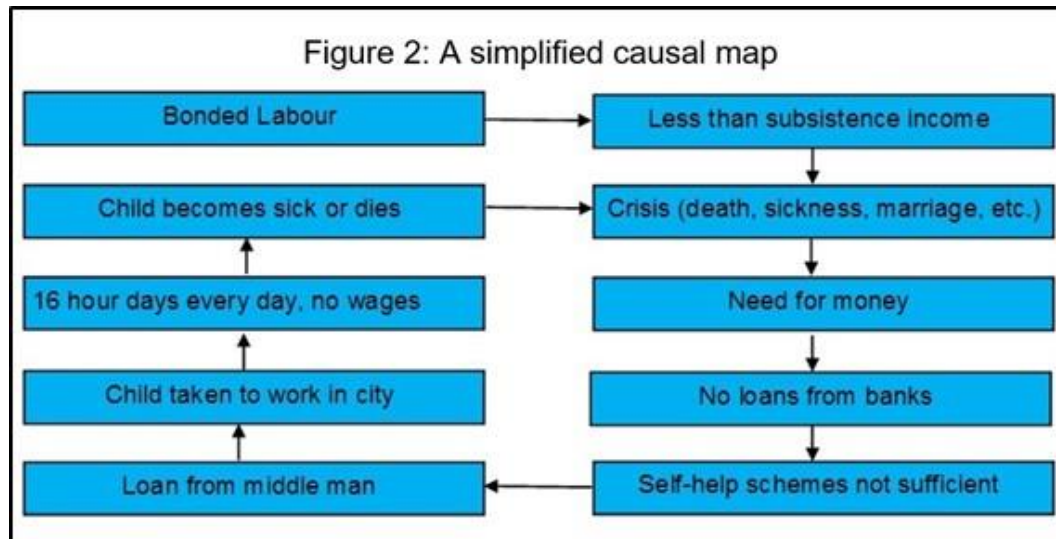
The vicious cycles and structural interactions between poverty and ill-health are well known (Aslan 2007, WHO 2002, Farmer 2001, Scheper-Hughes N. 1993). Although in both Bihar and Uttar Pradesh the government has made the improvement of health a priority and is investing in the health systems, this is somehow not benefitting these very poor and marginalized communities. Most of them belonging to lower castes, and face well known stigma and exclusion in accessing services. When NGO participants were asked to indicate the issues on which their interventions were focused, we found three strong clusters of activity around (a) supporting collective action, (b) awareness raising and (c) linking people to public services. There were almost no interventions related to health and none related to health loans. The programme thus provided both the NGOs and the funder with insights as to the potential value in terms of their primary objective of interventions that could improve the health status of these communities. If ill-health could be shown to increase the risk of slavery for the overall communities with which they were concerned, it would imply that improvements in health status should be a priority in slavery eradication efforts. To verify the analysis of the life stories and understand if health loans and poor access to health services were widespread in the lives of the people in the targeted communities, the researchers used participatory statistics.

Participatory statistics has the potential to satisfy a key requirement for impact evaluation and scaling up: the ability to generate quantitative data that can endorse attribution - thereby responding to the accountability and learning processes that are important for political and social sustainability. In the current example, an open-ended bottom up approach to the development of indicators and definitions, based on the multiple narratives described above, was used to generate data which was then subjected to formal statistical analysis.

A baseline study was conducted which involved 3,500 chulas (households) in 87 villages. Randomly selected programme beneficiaries collected the data and analysed the results of the prevalence to take collective action based on the results. For example one group saw that dowry was the second reason after health why people get into debts that lead to bonded labour

and asked the NGO to pilot a program with free marriage parties for couples who marry without a dowry. Within months a few couples did get married without a dowry. This approach allowed us to scale up decentralized, open-ended and action-oriented participatory processes involving both beneficiaries and NGOs and generate the data for a survey across the programme with sufficient statistical power. A follow-up survey will be carried out two and a half years after the baseline (Oosterhoff et al. 2016).

The preliminary analysis of 10 NGOs in Bihar showed that slavery was indeed widespread. Around 46% of households had a member in slavery. Almost 70% of people had no access to state health services, both slaves and non-slaves, and the vast majority of loans, almost 60%, were for healthcare. Slaves had more loans than non-slaves (85% vs 57%). These results suggest that health problems and health expenses are significant factors in terms of increasing the risk of slavery in spite of investments in the health system and the promotion of universal health care. A better understanding of the nature of the health problems (accidents, acute illness, chronic illness), health seeking behaviour, including the use of private or informal providers, and health expenditures is needed. The findings suggested that continuing to 'up-scale' the existing health system strategies was unlikely to improve the health status of those in the targeted communities.

6.   Conclusion

The evolution of participatory methods has accelerated rapidly over the past 20 years. Many participatory methods are now available which can genuinely engage stakeholders in the research process. These approaches can be used alone but as we can see they can also be effectively integrated with more traditional approaches such as statistical analysis. Because there are now effective methods to take them to scale they can be more systematically embedded in health policy design and programme scale up and learning strategies. Participatory Research is crucial when operating at the interface of medical and social systems - where norms, culture, power, resources etc. become critical variables in determining what is appropriate, relevant, effective and implementable.

References

Alsan, Marcella M., Michael Westerhaus, Michael Herce, Koji Nakashima, and Paul E. Farmer (2011). Poverty, Global Health and Infectious Disease: Lessons from Haiti and Rwanda. Infect Disease Clinics North America 25(3): 611-622.
www.ncbi.nlm.nih.gov/pmc/articles/PMC3168775/pdf/nihms301453.pdf

Annett, Hugh and Susan B. Rifkin (1995). Guidelines for rapid participatory appraisals to assess community health needs: a focus on health improvements for low-income urban and rural areas. WHO/SHS/DHS/98.5. Geneva: World Health Organization.
http://apps.who.int/iris/bitstream/10665/59366/1/WHO_SHS_DHS_95.8.pdf

Barker, Tom (2013). ORIE Research Summary 2. ORIE.
http://opendocs.ids.ac.uk/opendocs/bitstream/123456789/3182/1/ORIE_RSummary2_Stake Map_Web.pdf

Bowen JR: On the Political Construction of Tradition: Gotong Royong in Indonesia. The Journal of Asian Studies 1986, 45 (3): 545-561

Burns, D. and Worsley, S. (2015). *Navigating Complexity in International Development: Facilitating sustainable change at scale*. Rugby: Practical Action Publishing.
http://www.ids.ac.uk/publication/navigating-complexity-in-international-development-facilitating-sustainable-change-at-scale

Burns, D. (2014a). Systemic Action Research: Changing System Dynamics to Support Sustainable Change. *Action Research* 12(1):3-18.
https://www.researchgate.net/publication/274538278_Systemic_Action_Research_Changing _System_Dynamics_to_Support_Sustainable_Change

Burns, D. (2014b). Assessing Impact in Dynamic and Complex Environments: Systemic Action Research and Participatory Systemic Inquiry. *Practice Paper.* Centre for Development Impact.
http://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/4387/CDIPP8%20Assessin g%20Impact%20in%20Dynamic%20and%20Complex%20Environments%20Systemic%20A ction%20Research%20and%20Participatory%20Systemic%20Inquiry.pdf?sequence=1

Burns, D. (2007). *Systemic Action Research: A strategy for whole system change*. Bristol: Policy Press. http://press.uchicago.edu/ucp/books/book/distributed/S/bo13439973.html

Burns, D. and Evans, D. (2006). Effective health scrutiny: Learning from an action inquiry into health scrutiny of chronic disease management in six local authority areas. University of the West of England.

Burns, D. (1990) Collaboration in care: Integrating social and health care in North Staffordshire. Bristol University: School for Advanced Urban Studies.

Burns, D., Buttanshaw, C., Goicoechea, J. and Warren, G. (1999). Living Health in Proceedings of the 10th anniversary conference of the tipping the balance towards primary health care network, 88-104, Stockholm: Nordic School of Public Health.

Burns, D., Evans, D., Orme, J., Gray, S. and Weil, S. (2004a). Public health leadership in the real world: An action inquiry approach. In: From Evidence to Health: Eighth South West Public Health Scientific Conference, Weston-Super-Mare, UK, 12th May, 2004.
http://eprints.uwe.ac.uk/1989.

Burns, D., Evans, D., Orme, J., Gray, S. and Weil, S. (2004b). Leadership development through action inquiry. In *Sustaining Public Health in a Changing World: Vision to Action - 12th Annual Public Health Forum*, Brighton Centre, Brighton, UK, 22nd April, 2004. Available from: http://eprints.uwe.ac.uk/1988

Chambers R. Who counts? The quiet revolution of participation and numbers (1995). IDS Working Paper 296, IDS, Brighton

Chambers, R. (1981). Rapid rural appraisal: rationale and repertoire. Public Administration and Development 1:95-10 6. http://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/865/rc291.pdf?sequence=1

Cooke B, Kothari U (2001). The Case for Participation as Tyrrany. In Cooke B, Kothari U (eds.). *Participation: The New Tyranny?* London: Zed Books. http://isites.harvard.edu/fs/docs/icb.topic793411.files/Wk%209_Oct%2029th/Cooke_Kothari_2001_Participation_as_Tyranny.pdf

Cornwall, A. and Jewkes, R (1995) What is Participatory Research? *Social Science and Medicine* 41(12):1667-1676. http://www.civitas.edu.pl/pub/nasza_uczelnia/projekty_badawcze/Taylor/what_is_participatory_research.pdf

Easpaig, Bróna Nic Giolla (2015). Capturing collective processes of analysis in participatory research: An example from a memory work investigation into how gender and sexual identities are experienced. *International Journal of Social Research Methodology* . http://dx.doi.org/10.1080/13645579.2015.1111585

Emch, Root E D, Giebultowicz S, Ali M, Perez-Heydrich C, Yunus M (2012). Integration of Spatial and Social Network Analysis in Disease Transmission Studies. *Annals of the Association of American Geographers* 105(5):1004-1015. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806718/

EQUINET: TARSC and ALAMES (2014) Participatory action research in people centred health systems International Workshop Report: Cape Town, South Africa, 4 October 2014.

Evans, D., Pilkington, P. and McEachran, M. (2009) A systematic review of the impact of participatory approaches by UK public health units on health and social outcomes. In: *UKPHA 17th Annual Forum*, Brighton, UK, March 2009. Available from:http://eprints.uwe.ac.uk/526

Farmer P. (1999). *Infections and Inequalities: The Modern Plagues*. Berkeley: University of California Press, revised 2001 edition. www.ucpress.edu/book.php?isbn=9780520229136

Holland, J. (2013). *Who counts: The power of participatory statistics*. Rugby: Practical Action

Igras, Susan, Mariam Diakité & Rebecka Lundgren (2016). Moving from theory to practice: A participatory social network mapping approach to address unmet need for family planning in Benin. *Global Public Health*, DOI: 10.1080/17441692.2016.1147589. http://www.tandfonline.com/doi/pdf/10.1080/17441692.2016.1147589

Kleinman, A. Eisenberg L. Good B. (1978). Culture, Illness, and Care: Clinical Lessons from Anthropologic and Cross-Cultural Research. *Annals of Internal Medicine*. 88(2):251-258. doi:10.7326/0003-4819-88-2-251. http://annals.org/article.aspx?articleid=691789

Lindegger, G., C. Milford, C. Slack, M. Quayle, X. Xaba, and E. Vardas (2006). Beyond the Checklist: Assessing Understanding for HIV Vaccine Trial Participation in South Africa. *Journal of Acquired Immune Deficiency Syndrome* 43(5):560-566. http://www.saavi.org.za/beyond.pdf

Loewenson, R.; Laurell, A.C.; Hogstedt, C.; D'ambrouso, L. and Shroff, Z. (2014) *Participatory Action Research in Health Systems: A methods reader*, IDRC/ WHO www.equinetafrica.org/sites/default/files/uploads/documents/PAR Methods Reader2014 for web.pdf

Mansuri, Ghazala and Vijayendra Rao (2003). Evaluating Community-Based and Community-Driven Development: A Critical Review of the Evidence. Development Research Group. The World Bank. http://siteresources.worldbank.org/INTECAREGTOPCOMDRIDEV/Resources/DECstudy.pdf

Miles, D., Cronin, D., Tolley, F., Lewendon, G., Chapel, H., Mackenzie, I., Howard, J., Morgan, K.,Kendall, L., Brown, S., Burns, D., Evans, D., Orme, J., Gray, S. and Weil, S. (2006) Public health leadership in the real world: The role of the director of public health. Public Health Medicine, 6 (2). pp. 57-9. Available from: http://eprints.uwe.ac.uk/4638

Mohanan M, Vera-Hernández M, Das V, et al. (2015). The Know-Do Gap in Quality of Health Care for Childhood Diarrhea and Pneumonia in Rural India. JAMA Pediatrics 169(4):349-357. http://archpedi.jamanetwork.com/article.aspx?articleid=2118580

Needle R, Trotter R, Singer M, Bates C, Page J, Metzger D, Marcellin LH (2003). Rapid assessment of the HIV/AIDS crisis in racial and ethnic minority communities: an approach for timely community interventions. *American Journal of Public Health* 93(6): 970-979, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447879/pdf/0930970.pdf

Ngaruko, D D (2015). Tanzania: A Hierarchical Cluster Analysis Approach. Huria Journal 20:114-132. http://www.out.ac.tz/files/documents/HURIA JOURNAL VOL. 20 2015.pdf#page=118

Oosterhoff P., Bharadwaj S., Burns D., Mohan Raj A., Nanda R. B., Narayanan P. (2016). Using Participatory Statistics to Examine the Impact of Interventions to Eradicate Slavery: Lessons from the Field. *CDI Practice Paper 16*. http://www.ids.ac.uk/publication/using-participatory-statistics-to-examine-the-impact-of-interventions-to-eradicate-slavery-lessons-from-the-field

Oosterhoff, P., Saprii, L., Kharlyngdoh, D. and Albert, S. (2015a). When the Hen Crows: Obstacles that Prevent Indigenous Women from Influencing Healthcare Policies - A Case Study of Shillong, Meghalaya, India. IDS Evidence Report 1119. http://www.ids.ac.uk/publication/when-the-hen-crows-obstacles-that-prevent-indigenous-women-from-influencing-health-care-policies-a-case-study-of-shillong-meghalaya-india

Oosterhoff P., Dkhar B., Albert S. (2015b) Understanding unmet contraceptive needs among rural Khasi men and women in Meghalaya. Culture, Health and Sexuality 17(9). https://www.researchgate.net/publication/279304340_Understanding_unmet_contraceptive_needs_among_rural_Khasi_men_and_women_in_Meghalaya

Oosterhoff P. (2009). Observations on action research with HIV-positive women and state service providers in northern Vietnam. Medische Antropologie 21(2):257-275. http://s3.amazonaws.com/academia.edu.documents/44285961/Observations_on_action_research_Vietnam_Oosterhoff.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1

469874572&Signature=eBj1OMXFAeDKdGRCqHtMfJhNh8U%3D&response-content-disposition=inline%3B%20filename%3DObservations_on_action_research_with_HIV.pdf

Oosterhoff P., Hardon A., Nguyen T. A., Pham N. Y., and Wright P. (2008). Dealing with a positive result: risks and responses in routine HIV testing among pregnant women in Vietnam. *AIDS care* 20(6):654-659. https://hal.archives-ouvertes.fr/hal-00513431/document

Oosterhoff P., White J., Nguyen Thi Huong (2011). Family health consequences of modernisation programmes in Black Thai communities. Culture Health and Sexuality 13 Supplement 2:S217-33. https://www.researchgate.net/publication/51049633_Family_health_consequences_of_modernisation_programmes_in_Black_Thai_communities

Percy-Smith, B. (2007) 'You think you know? … You have no idea': youth participation in health policy development. Health Education Research; 22 (6) pp 879-894. http://her.oxfordjournals.org/content/22/6/879.full.pdf+html

Percy-Smith, B., Burns, D., Weil, S. and Walsh, D. (2003). *Mind the gap: healthy futures for young people in Hounslow*. University of the West of England and Hounslow Community Health Council. ISBN 1860433588.

Riemenschneider, Barca and Holland (2013). Scoring perceptions of services in the Maldives: Instant feedback and the power of increased local engagement. In *Who Counts: the Power of Participatory Statistics*. Practical Action Publishing. http://www.intrac.org/data/files/Oxford_Policy_Management_-_final.pdf

Scheper-Hughes N. 1993 *Death without Weeping: The Violence of Everyday Life in Brazil.* Berkeley: University of California Press. (Second edition, paperback). http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X1992000200018

Schmittdiel, J. A., Grumbach, K., & Selby, J. V. (2010). System-Based Participatory Research in Health Care: An Approach for Sustainable Translational Research and Quality Improvement. *Annals of Family Medicine* 8(3):256-259. http://www.annfammed.org/content/8/3/256.long

Sharp, C; Burns, D; & Bass, L. (2008). After the floods: the lessons learnt. London: British Red Cross.

SLF (2015). Community Participation in Action for the Prevention of TB and HIV in South Africa - Photo Voice. www.youtube.com/watch?v=aoNSHa7aDYU

Tatem, Andrew J, Susana Adamo, Nita Bharti, Clara R Burgert, Marcia Castro, Audrey Dorelien, Gunter Fink, Catherine Linard, Mendelsohn John, Livia Montana, Mark R Montgomery, Andrew Nelson, Abdisalan M Noor, Deepa Pindolia, Greg Yetman and Deborah Balk (2012). Mapping populations at risk: improving spatial demographic data for infectious disease modelling and metric derivation. *Population Health Metrics* 10:8. https://pophealthmetrics.biomedcentral.com/articles/10.1186/1478-7954-10-8

UNAIDS and Royal Tropical Institute (2004). *Techniques and Practices for Local Responses to HIV/AIDS*. Royal Tropical Institute. www.bibalex.org/Search4Dev/files/281403/113106.pdf

UNHCR WASH Manual - Preliminary Version, 2015. UNHCR http://wash.unhcr.org/download/unhcr-wash-manual-unhcr-2015-2/

UNHCR, WHO 2008 Rapid Assessment of Alcohol and Other Substance Use in Conflict-affected and Displaced Populations. UNHCR, WHO.
http://www.who.int/mental_health/emergencies/unhcr_alc_rapid_assessment.pdf

Vallenga, Dineke, Mieke H. F. Grypdonck, Lucienne J. R. Hoogwerf and Francis I. Y. Tan (2009). Action research: what, why and how ? *Acta Neurologica Belgica* 109:81-90.
http://www.actaneurologica.be/acta/download/2009-2/02-Vallenga%20et%20al.pdf

World Health Organization (2002). Dying for change — poor people's experience of health and ill health. Geneva: WHO; 2002. http://www.who.int/hdp/publications/dying_change.pdf

Chapter 11: Incorporating gender analysis into health systems implementation research

**Rosemary Morgan, Department of International Health,
Johns Hopkins Bloomberg School of Public Health
Sally Theobald, Liverpool School of Tropical Medicine
Kate Hawkins, Pamoja Communications
Linda Waldman, Institute of Development Studies
Helen Elsey, Leeds Institute of Health Sciences, University of Leeds**

1.   Introduction

Implementation research studies the implementation of interventions to promote the successful uptake and scale-up of evidence-based policies, programmes, and practices (collectively called interventions) (Sanders & Haines 2006; Peters, Adam et al. 2013; Bhattacharyya et al. 2009). Implementation research helps us to understand how and why interventions fail or work in real-world settings by exploring what factors positively and negatively affect implementation in specific contexts, in addition to finding solutions to improve overall implementation (Peters, Tran, et al. 2013; Peters, Adam, et al. 2013). When considering context within implementation research it is important to explore how cultural and social factors affect the implementation of an intervention. Such factors include the consideration of gendered power relations and their role in perpetuating vulnerability and marginalization within and outside the health system (Östlin et al. 2006; Ravindran & Kelkar-Khambete 2007).

This chapter explores how gender analysis can be incorporated into health systems implementation research. This is the process of analysing how gendered power relations influence the implementation of an intervention, as well as the extent to which the research process itself progressively transforms gendered power relations, or at least does not exacerbate them. Gender is defined as the "socially constructed roles, behaviours, activities, and attributes that a given society considers appropriate for men and women" (WHO 2016). Gender is different from sex, which refers to the biological chromosomal characteristics that distinguish males, females, and intersex people (Sen et al. 2007). The meaning of gender can therefore extend beyond men and women to include a range of gendered identities while also varying over time and across contexts. As well as analysing differences between men, women, and people of other genders, gender analysis also explores differences among these categories. It takes an intersectional approach which examines "*gender in relation to other social stratifiers, such as class, race, education, ethnicity, age, geographic location, (dis)ability and sexuality, etc., how these markers dynamically interact, [and] how power plays out at multiple levels and through diverse pathways to frame how vulnerabilities are experienced*" (Morgan et al. 2016: 2; Larson et al. 2016).

Within implementation research, there is a clear need to understand context and the ways in which gender, power and other social stratifiers shape systems', individuals' or households' abilities to access and use interventions. There are many examples of effective interventions that are not successfully implemented, or not implemented at all. When interventions are implemented, it is not a given that they will be utilized as intended. For example, while sanitation can prevent a host of communicable and non-communicable diseases in low- and middle-income countries (LMICs), in some contexts latrines are not used despite their availability. In their study of latrine use in Zambia, Thys et al. (2015) found that taboos associated with in-laws and sexually mature daughters meant men did not use latrines. In situations like this, implementation research that carefully considers gender dynamics can be

used to help health systems researchers, communities, policy makers, and practitioners understand the factors affecting the adoption, uptake, and/or use of an intervention.

When designing and implementing health systems interventions, there is often the assumption that an intervention will be equally effective for men, women, and people of other genders across all socio-economic and ethnic strata (Östlin et al. 2006). Likewise, implementers often fail to recognize how gendered power relations can affect how someone interacts with, accesses, uses, or generally responds to an intervention. For example, most health systems research, including implementation research, uses gender-neutral expressions and words, such as 'patients', 'health care providers', 'adolescents', 'children', or 'employees', failing to make either the sex (or other social markers) of the participants explicit (Östlin et al. 2006). Failure to disaggregate health systems data by sex masks differences between males and females and failure to consider gender obscures people of other genders, overlooking the roles that gendered power relations play in creating different health systems' needs, experiences and outcomes, including how they affect implementation (Morgan et al. 2016).

This chapter outlines how gender analysis can be incorporated into health system implementation research *content, process,* and *outcomes.* While each is discussed separately, we recognize that "*they interact, overlap and reinforce one another; and an approach that takes forward gender within these areas is mutually reinforcing*" (Morgan et al. 2016:2). While it is important that genders other than men and women are considered within such analyses, we note that the majority of examples used within this chapter deal primarily with relations between men and women, and that the body of literature on implementation research, and health systems more broadly, is inadequate when it comes to other genders.

2.   Gender analysis within implementation research content

Incorporating gender analysis into implementation research content requires us to explore how gendered power relations affect the implementation of an intervention. In order to do so, implementation research needs to first be disaggregated by sex and other social stratifiers. Alongside data disaggregation, implementation researchers can use gender frameworks and gender analysis questions to explore the role of gender power relations in relation to implementation.

Disaggregating data by sex and other social stratifiers
Gender analysis cannot be incorporated into implementation research without first disaggregating data by sex. Sex disaggregation entails differentiating between males and females during data collection, and ensuring that this information is recorded and maintained (Morgan et al. 2016; Nowatzki & Grant 2011). Research which fails to disaggregate data by sex can miss important differences between men and women, such as how gendered power relations shape men's and women's experiences and outcomes in relation to the implementation of an intervention. Even when sex disaggregation does occur, most health systems research assumes that there are two genders, men and women, and fails to enquire about gender diversity. As a result, it does not consider how people of other genders experience various health interventions.

For example, implementation research which does not report sex as a variable and instead uses gender-neutral terms such as 'community member' or 'patient', may fail to see how women's lack of financial resources or decision-making power within the home can affect their use of an intervention. Likewise, an intervention whose success requires lower-tiered health workers, such as community health workers, to change their working practices (e.g. work late), is likely to have additional ramifications if it does not consider the gendered distribution of the health workforce and the fact that women are usually employed in lower-tiered occupations (George 2008). Such a requirement not only puts strain on female health workers who have

additional caring responsibilities, impacting upon their relationships within their homes, it also places these women at risk if they are required to journey home after dark.

Where possible data should also be disaggregated by other social stratifiers, such as age, race, class, ethnicity, geographic location and (dis)ability (Larson et al. 2016). The impact of an intervention on adolescent girls, for example, may be different to the impact experienced by older women; similarly, there may be different factors affecting how men and women interact with an intervention in rural areas compared to urban areas, such as the need for transportation and other resources to reach a distant health facility.

Much implementation research in low and middle-income countries utilizes routine Health Management Information System (HMIS) data to get a picture of the overall health landscape. Such data may already disaggregate information by male and female (although this is not always the case), and it is unlikely that any other stratification will be available. Where possible, data should be disaggregated by strata relevant to the context; existing evidence reviews and informative primary qualitative data collection could inform decisions about which strata are appropriate for a particular context. If resources are limited, however, researchers should consider conducting smaller qualitative studies or questionnaires to explore differences by gender and other social stratifiers. Incorporating a gender and/or intersectional lens into implementation research will therefore often require primary data collection, which has additional resource implications.

Example: Implementation research on tuberculosis in Ethiopia: insights from gender analysis

The implementation problem set in context: Tuberculosis (TB) is one of the major causes of morbidity and mortality in Ethiopia. TB Control Programmes rely on passive case finding to detect cases, and TB notification remains low in Ethiopia despite major expansion of health services. Rural populations with high levels of poverty and gender inequity are most likely to have unmet health needs and undiagnosed TB cases. These vulnerable communities include people living in rural and remote settings with limited access to TB diagnostic facilities due to lack of awareness, socio-cultural and gender-related barriers, TB related stigma, and inability to afford for the time and expenses related to seeking diagnosis and treatment.

The implementation research approach: The aim was to ensure the effective implementation of a TB control intervention within these vulnerable communities. Ethiopia has established a Health Extension Program (HEP) which includes the training and deployment of female health extension workers (HEWs) based in local communities to improve access to primary health services. A community-based intervention package was implemented in Sidama zone, Ethiopia in partnership between the Sidama Health zone and researchers at REACH Ethiopia, Liverpool School of Tropical Medicine and the Global Fund with funding from TB REACH. The package included advocacy, training, engaging stakeholders and communities and active case finding by women HEWs at village level. HEWs conducted house-to-house visits, identified individuals with a cough for two or more weeks, with or without other symptoms, collected sputum, prepared smears and supervised treatment. Supervisors transported smears for microscopy, started treatment, screened contacts and initiated Isoniazid preventive therapy (IPT) for children. Ongoing process evaluation involved multiple methods: outcomes were compared with the pre-implementation period and a control zone; complimentary qualitative research (interviews and Focus Group Discussions) were conducted to understand community and provider perceptions and experiences.

Taking a gender perspective: All outcome data was gender disaggregated. Between October 2010 and December 2011 (at the beginning of the process), HEWs identified

49,857 individuals (29,314 [60%] women) with cough for two or more weeks, with or without other symptoms. Of these, 2,262 (1,199 [53%] women) were smear-positive (PTB+). The male to female ratio among PTB+ cases changed from 1.3:1 before the intervention to almost 1:1. The proportion of women among PTB+ cases was lower in the public health facilities than in the community (44% and 53%; P=0.001). Some participants indicated they would have been unable to get a diagnosis without the intervention due to direct and opportunity costs, and would have instead "waited at home for death". Respondents often referred to multiple barriers to diagnosis faced pre-intervention. For example, distance was particularly challenging for women, the poor, elderly and the very sick:

"I am not able to go to far places to be treated because I don't have money for transportation and food. Here in my community, without going to the health centre, I am getting treatment…It is what makes me very happy" (Pt, Woman, 49yrs)

Community-based treatment reduced difficulties associated with adherence, although lack of food remained an important issue for some patients. The interventions have reduced barriers to services with poor women who had previously faced difficulties travelling to health centres. The proportions of children and elderly among symptomatic and PTB+ cases also increased during the implementation period, and these are also vulnerable groups better reached by an intervention package that is embedded in the community. Qualitative assessments with female HEWs showed that providers described commitment or "devotion" to improving the health of their communities who lacked education on health matters, yet accepted guidance through the community engagement activities, and highlighted the package improved access and awareness, particularly for the very poor and women. HEWs felt job satisfaction collecting and preparing smears, the preventive and curative aspects of their work and felt guided and supported by supervisors. Being a HEW involved in "TB work" warranted "respect" from the community.

The implementation research highlights the importance of sex-disaggregated data to assess changes of the package at the community level, as well as ongoing gender aware qualitative assessments with key providers (in this case all female HEWs) to assess the impact on their experience and workload.

Adapted from Yassin et al. (2013); Datiko et al. 2015; Tulloch et al. 2015; with thanks to Mohammed Yassin, Daniel Gemechu Datiko, Luis Cuevas and Olivia Tulloch.

Disaggregating data by sex and other social stratifiers alone, however, does not constitute gender analysis (Morgan et al. 2016). Incorporating gender analysis into implementation research goes beyond disaggregating data by sex and other social stratifiers to consider how gender power relations impact upon how different categories (rural men, old women, urban teenagers) use, interact with, and respond to an intervention and its implementation; and in turn to consider how an intervention and its implementation may impact upon gender power relations. By identifying which (and how) gendered power relations impact upon an implementation, implementers can amend their implementation strategies accordingly. In order to do so, researchers can use gender frameworks and gender analysis questions to interrogate disaggregated data.

Using gender frameworks within implementation research
Gender frameworks can be used as a guide to help researchers structure their thinking, research questions, data collections tools, and analysis (Morgan et al. 2016). They can help researchers think about what aspects of gendered power relations may affect an intervention and its implementation, and incorporate specific gender analysis questions related to these aspects into their data collection tools and analysis. There are a number of gender frameworks

that can be used within implementation research; Table 1 outlines a few frameworks specifically related to health and/or health systems.

Table 1: Gender Frameworks that Address Health Systems

Specific Frameworks:
- Gender Analysis Toolkit for Health Systems (JHPIEGO 2016)
- Guide for analysis and monitoring of gender equity in health policies (PAHO 2009)
- Addressing Gender and Women's Empowerment in mHealth for MNCH: An analytical Framework (Deshmukh & Mechael 2013)
- Guidelines for the Analysis of Gender and Health (LSTM 1996)
-

A framework Summary:
- Ten Gender Analysis Frameworks for Health Systems Research (RinGs 2015)

The gender frameworks above can help researchers think about and develop questions related to key domains that constitute gendered power relations. The framework in Table 2 below, originally presented in Morgan et al. (2016), organizes these domains into four categories: *who has what* (access to resources); *who does what* (the division of labour and everyday practices); *how values are defined* (social norms, ideologies, beliefs, and perceptions) and *who decides* (rules and decision-making). The framework also demonstrates how these domains are not static, and are instead negotiated by people and their environments, changing over time and across contexts.

Table 2: Gender Analysis Framework: Gender as power relation and driver of inequality

What constitutes gendered power relations?	
Who has what?	Access to resources (education, information, skills, income, employment, services, benefits, time, space, social capital etc.)
Who does what?	Division of labour within and beyond the household and everyday practices
How are values defined?	Social norms, ideologies, beliefs and perceptions
Who decides?	Rules and decision-making (both formal and informal)
How power is negotiated and changed?	
Individual/ People	Critical consciousness, acknowledgement/ lack of acknowledgement, agency/apathy, interests, historical and lived experiences, resistance or violence
Structural/ Environment	Legal and policy status, institutionalisation within planning and programs, funding, accountability mechanisms

Incorporating gender analysis questions into implementation research

The purpose of implementation research is to improve and support the implementation of an intervention, and ensure its successful outcome. Incorporating gender analysis questions into implementation research helps researchers explore how gendered power relations contribute to the success or failure of implementation (i.e. how gendered power relations affect whether or not an intervention is used, adopted, suitable, etc.). By incorporating gender analysis questions into the data collection and analysis process, researchers can begin to understand which gendered power relations affect how different groups use, interact with, and respond to an intervention, in addition to better understanding the positionality and motivations of those involved in the research process and intervention. These gendered power relations impact upon whether or not implementation is successful; and importantly, on whether the intervention itself has the potential to exacerbate, maintain or transform gendered power relationships at different levels - within households, communities and institutions.

While an intervention might work well, for example, it can serve to exacerbate challenging gender norms. An intervention which relies on women's unpaid labour, for example, can exacerbate gender relations which undervalue women's work compared to men's. Alternatively, an HIV intervention which mandates HIV testing may increase the number of people who know their HIV status and be considered a positive public health intervention by health system actors. However, for populations who are marginalized by discriminatory laws or social stigma - such as homosexuals, people living with HIV, and sex workers - such interventions will be experienced as abusive and marginalizing and may, in turn, reduce the numbers of people willing to use health services to help them manage their HIV status in a positive and healthy way. Such mandatory testing can also lead to gender-based violence and/or social rejection, particularly for women who find out their HIV positive status before their partners are tested, and are then blamed for bringing HIV into the relationship (Jewkes et al. 2003).

Peters, Tran, et al (2013) present a framework to help implementation researchers evaluate the success or failure of implementation against different outcomes of implementation. Such characteristics are described as 'implementation outcome variables', and serve as indicators of how well implementation is actually working (i.e. whether or not it is achieving its desired results). The implementation outcome variables include: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, coverage and sustainability. A definition of each variable is provided in Table 3 below. The relative importance of each outcome variable will be dependent on the intervention that is being delivered and, as such, researchers may choose not to explore all outcome variables within their data.

In relation to these outcome variables, gender analysis questions could explore how gendered power relations affect whether or not an intervention is acceptable, adopted, appropriate, feasible, implemented as intended (fidelity), and sustainable. In Mozambique, for example, while policies recommended that women be recruited as paid community health workers (CHWs) due to their experience with newborn and child health, men were prioritized for these positions as they were viewed as the key breadwinners within the family and more in need of employment (Chilundo et al. 2015). These decisions limited the kinds of health care possible by CHWs as men were not seen as appropriate midwives and they themselves avoided intimate health care associated with pregnancy and childbirth. In this case, the gendered selection of paid CHWs by communities undermined the implementation of the policy, and, as a result, the policy did not achieve its intended outcomes of recruiting more women as paid CHWs and improving newborn and child health. If such factors were identified prior to implementation, implementers could have incorporated strategies within the community to support women's employment as CHWs - or to help men overcome the barriers of dealing with issues of pregnancy and childbirth - to help ensure the intervention was implemented effectively.

Using the gender framework presented in Table 2, Table 3 below provides illustrative gender analysis questions against each of the outcome variables listed above. These serve as a guide to show how the content and range of relevant gender questions will depend on the type of implementation research. Factors that influence gender and power are difficult to contain within neat discrete categories, and hence there are some overlaps between the factors in the gender power relations domain. As it may not always be possible to address all the questions laid out below, researchers should start by identifying important gender analysis questions which are relevant to their implementation research.

Table 3: Illustrative Gender Analysis Implementation Research Questions depending on content and focus of the implementation research

Implementation Outcome Variable: Acceptability - The perception among stakeholders that an intervention is agreeable	
Gender Power Relations Domain	**Illustrative Gender Analysis Questions**
Access to resources	To what extent do women's (frequent) lack of skills and resources (education, money, technology, employment) or autonomy affect whether or not others perceive their involvement in the intervention as acceptable?
Division of labour and everyday practices	Does men's and women's work inside and outside the home affect whether or not others perceive their involvement in the intervention as acceptable?
Social norms	How do social and cultural gender norms affect whether or not the intervention is accepted by the community, e.g. do cultural beliefs about women as child bearers and mothers influence their involvement in a family planning intervention?
	How do the conditions at health facilities affect access? To what extent do health facilities provide services with appropriate conditions (such as functioning toilets, bathing areas for inpatient facilities, shelter from sun/rain in the waiting area) and confidential services? Can women request a female health care provider if they wish to?
Rules and decision-making	Who decides whether or not it is acceptable for someone to participate in an intervention? How do they decide this? Are women or other marginalized populations (transgender people, ethnic minorities, migrants, inhabitants of informal settlements, people employed in illegal occupations, etc.) excluded?
Implementation Outcome Variable: Adoption - The intention, initial decision, or action to employ a new intervention (i.e. uptake)	
Gender Power Relations Domain	**Illustrative Gender Analysis Questions**
Access to resources	To what extent are marginalized populations able to access relevant information and care related to an intervention?
Division of labour and everyday practices	How do women's social roles, such as childcare, infant feeding and other reproductive tasks, affect their access to and utilization of an intervention?
Social norms	How does stigma and or access to resources inhibit certain men and women from accessing or using an intervention? Do interventions which are targeted at women, such as maternal and child health and family planning services, exclude men?
Rules and decision-making	Who decides whether and how much household resources should be used to pay for health care services? How might this affect an intervention? Do women require the permission of a male partner or relative to use the intervention?

Implementation Outcome Variable: Appropriateness - The perceived fit or relevance of the intervention in a particular setting or for a particular target audience or issue	
Gender Power Relations Domain	**Illustrative Gender Analysis Questions**
Access to resources	To what extent do women's (frequent) lack of skills and resources (education, money, technology, employment) affect whether or not others perceive their involvement in the intervention as relevant?
Division of labour and everyday practices	To what extent are the intervention activities, such as health outreach visits or clinics, organised considering men's and women's agricultural, economic, or caretaking activities in their communities? Does involvement in the intervention compromise any implementers' safety? Or bring additional tasks to certain groups that may be unpaid or unremunerated? (i.e. do they rely on the labour of women volunteers who have to travel after dark)?
Social norms	How do women and men within households and communities prioritise individuals' involvement in an intervention, e.g. is the intervention more likely to be seen as relevant for men due to their role as providers or as relevant for women because of its unpaid, low-prestige status? Does the implementation problem and design draw on health providers' (and others') tacit knowledge? Does it incorporate both male and female perspectives?
Rules and decision-making	Who decides whether or not someone can participate in an intervention - and at what level, i.e. within households, communities, institutions? And how is this decided?
Implementation Outcome Variable: Feasibility - The extent to which an intervention can be carried out in a particular setting or organization	
Gender Power Relations Domain	**Illustrative Gender Analysis Questions**
Access to resources	To what extent do women and men (or other marginalised categories of people) have the same access to educational and training opportunities? To what extent do family support and roles help or limit opportunities for training by gender, marital status, age, or other social stratifiers? How might this affect stakeholder engagement within an intervention? To what extent do women (or other marginalised categories) have sufficient literacy, autonomy, and access to technology to effectively use an intervention? To what extent is protective health equipment and gear made available and does it fit bodies that are not the male standard?
Division of labour and everyday practices	To what extent are women more or less likely to work in frontline service delivery in poorly-compensated (including volunteer) or less-supported positions than

	men? How does this affect who implements an intervention and how? How do men's and women's roles and responsibilities affect the use of products used within the intervention (e.g. bed nets, vaccinations)? What are the challenges different groups of women and men might face in adhering to long-term treatment (e.g. for tuberculosis, HIV or diabetes)? Are they appropriately supported, or stigmatised, within health systems and community based structures?
Social norms	How do women and men within households and communities prioritise individuals' access to medical technologies or commodities used within an intervention, e.g. are boys or girls more likely be prioritised for oral rehydration therapy (ORT)? How do social norms and notions of masculinity and femininity influence men's and women's decisions to use the protective equipment required in an intervention?
Rules and decision-making	To what extent does regulation stand in the way of making services used within the intervention more widely accessible for women or marginalised groups, e.g. medical abortion, family planning? What is the effectiveness of regulatory mechanisms to ensure that medical products for women or other marginalised groups are not misused, e.g. oxytocin to augment labour?

Implementation Outcome Variable: Fidelity - The degree to which an intervention was implemented as it was designed in an original protocol, plan, or policy.

Gender Power Relations Domain	Illustrative Gender Analysis Questions
Access to resources	To what extent have those in leadership positions received training in gender sensitivity or gender mainstreaming? To what extent does this training emphasis the need to proactively think about gender and power relations and how they may shape an intervention and exacerbate or minimize harm?
Division of labour and everyday practices	How might participation in an intervention affect health workers' relationships within the home? Will participation in an intervention compromise their safety? To what extent are there differences by gender and other social markers in participation, decision-making, and planning of interventions?
Social norms	Are female and male health providers recognised differently within an intervention? Do they have different needs? To what extent are female providers expected to provide more emotional support, or do more caring work than male providers? Are male providers expected to work in more dangerous contexts or travel longer distances?

Rules and decision-making	Has gender been mainstreamed into an intervention design, and if so how, and with what impact?

Implementation Outcome Variable: Implementation Cost - The incremental cost of the delivery strategy. The total cost of implementation also includes the costs of the intervention itself.

Gender Power Relations Domain	**Illustrative Gender Analysis Questions**
Access to resources	Do male and female implementers receive the same level of pay? Do male and female volunteers receive similar incentives? Do performance-based incentives mean the same thing for female and male health workers across and within cadres? How might this affect an intervention? Are services or goods which would increase men's or women's involvement included in the intervention included in the budget?
Division of labour and everyday practices	Are opportunity costs appropriately documented from different perspectives in cost calculations, e.g. the opportunity costs of seeking care/accessing an intervention (and not being able to participate in paid/unpaid work)? From an implementers' perspective how might costs of participating affect women and men differently?
Social norms	What are the social norms around negotiating for the prices of goods and services? Does having a male or female negotiator affect the cost?
Rules and decision-making	Who decides what to spend money on? How might this affect what is included within the budget?

Implementation Outcome Variable: Coverage - The degree to which the population that is eligible to benefit from an intervention actually receives it.

Gender Power Relations Domain	**Illustrative Gender Analysis Questions**
Access to resources	To what extent do user fees or the removal of user fees have an impact on women and other marginalised groups? Have disaggregated information on out-of-pocket expenditures on health for different groups been obtained? Does an intervention incur more out-of-pocket expenditures for men or women? And what is the impact of this on individuals and households? Who has access to the skills, devices and technology that transmits and processes health information? How do they use this information?
Division of labour and everyday practices	How might men or women's responsibilities both inside and outside the home affect their ability to participate in the intervention?
Social norms	Are health workers in public facilities more likely to respond to certain groups of clients based on perceived ability to pay, gender etc.? How might this affect an intervention?

Rules and decision-making	Are those with decision-making power included within the intervention? How might their lack of inclusion affect ability to access the target population?
Implementation Outcome Variable: Sustainability - The extent to which an intervention is maintained or institutionalized in a given setting.	
Gender Power Relations Domain	**Illustrative Gender Analysis Questions**
Access to resources	Who is more likely to have higher literacy levels and access to social capital enabling them to participate more effectively in health committees and other forms of health/ intervention planning?
Division of labour and everyday practices	To what extent are there differences by gender and other social markers in participation, decision-making and planning of interventions?
Social norms	Does an intervention encourage the participation of men in women's and children's health? If yes, how, and on what terms? Does it rely on women's unpaid labour?
Rules and decision-making	To what extent do policies exist to ensure that women are represented on decision-making bodies related to an intervention?

* Working definitions of Implementation Outcome Variables from Peters, Tran, et al (2013)

Incorporating gender analysis into data collection processes

Incorporating gender analysis into implementation research also includes understanding how gendered power relations can affect the data collection process (i.e. how and where data collection occurs and who is involved). This is an important consideration for implementation research as gendered power relations can negatively affect the type and quality of data that is collected. Gendered power relations can, for example, influence the accuracy and validity of data collection, which ultimately affects the results and recommendations reported and impacts upon the overall success of the intervention. When conducting implementation research it is important to think about how gender as a power relation influences: who participates within the research (as respondents, data collectors, and data analyzers), who is present during data collection, and when data is collected and where (Theobald et al. 2006; Morgan et al. 2016).

Who is involved within the research process?

Within implementation research, it is important to consider who your respondents are and who might be excluded. Where possible, data should be collected from all relevant stakeholders and any categories of people excluded from participation due to gender power relations should be identified. Implementation research involving interventions focusing on maternal and child health, for example, may exclude men due to social norms that dictate that such issues are the responsibility of women. By excluding men, researchers may fail to recognize men's decision-making role within the household and their influence over the health of their wives, daughters, or daughters-in-law (Morgan et al. 2016; Thapa & Niehof 2013). Similarly, women may be excluded from research due to lower levels of literacy or education, because they may require additional permission to travel to research locations, or because they have less leisure and privacy (Morgan et al. 2016).

When collecting data, implementation researchers also need to consider how the gender of the data collectors influences the accuracy and quality of the data collected. In some contexts, it may be inappropriate for male or female data collectors to record information from someone of the opposite sex. It is equally important to think about how other characteristics of data collectors, such as age, ethnicity, occupation, or class, may affect the quality and accuracy of the data collection (Morgan et al. 2016). Data collectors' gender can also affect access (e.g.

whether or not someone is allowed entry into a home or access to children). Depending on the intervention being implemented, in some contexts female data collectors, for example, are more likely to be provided access, particularly if a woman's husband or father-in-law is not home. Conversely, female data collectors may have difficulty getting access males when the implementation research involves male sexual health and behaviour.

In addition, researchers' own underlying gender biases and assumptions can affect the quality and accuracy of the data collected, data analysis, and the results reported. Adequate training and supervision is therefore needed to help researchers recognize their own potential gender biases. This should be accompanied with processes that support reflection on data collection and analysis, such as joint reviews or debriefing meetings (Morgan et al. 2016).

Who is present during data collection?

It is important for implementation researchers to consider who is present during the data collection process. For example, the quality and accuracy of data may be affected if both men and women are present during interviews, focus group discussions, or surveys, as each may be reluctant to share information if someone from the opposite sex is present (Hunt 2004; Morgan et al. 2016). Alternatively, in healthcare settings, a patient may be reluctant to provide sensitive information if a healthcare worker of the same or opposite sex is present, or healthcare workers may be unwilling to speak up if a female/male superior is present (Morgan et al. 2016).

When data is collected and where?

The role of gendered power relations in relation to when data is collected, and where, also needs to be considered. The timing and location of data collection, for example, can negatively affect people's involvement in the research project. Due to women's and men's different responsibilities in relation to work and family life, they may be available at different times of the day. Women, for example, often have a double-burden in relation to work and home life, which may affect their ability to participate. Similarly, men may be working away from home and unable to participate during the day or on weekdays. It is therefore important to choose a convenient time and place to carry out data collection in order to ensure that relevant individuals are not excluded (Hunt 2004; Morgan et al. 2016).

Reducing gender bias within the research process is an important component of incorporating gender analysis within implementation research. Without consideration of gender within the research process, the overall quality and accuracy of the data could be affected, along with the outcomes of the intervention itself. Incorporating gender into the research process therefore has the potential to produce higher equality and more effective implementation research.

> **Example: Addressing gender and power to in the development of HIV interventions to better meet the needs of needs of sex workers in India**
>
> **The implementation problem set in context:** HIV is a priority area for action in India. Gender, power and social exclusion shape both who is vulnerable to HIV infection and ability to access and adhere to quality care and treatment. Certain group, such as sex workers, are particularly vulnerable to HIV and are considered a key population group that should be prioritised in the HIV response. Sex workers in many Indian contexts are stigmatized, face extreme discrimination, are not organized in groups and experience violence at the hands of police and family members.
>
> **The implementation research approach**: Implementation research has been ongoing to develop the most appropriate strategies to provide HIV and AIDS services for female sex workers. The government of India developed the National AIDS Control Program (NACP), which involves the implementation of targeted interventions to reduce HIV for

groups considered at "high risk", including sex workers. A qualitative process evaluation was undertaken in two states: Andhra Pradesh and Karnataka, to assess the ways in which targeted interventions are appropriately adapted to sex workers' needs and the changing contextual and programmatic factors.

Establishing outreach activities for sex workers, many of whom were female and illiterate was challenging; implementers needed to be aware of how gendered power shapes their experiences. Following ongoing dialogue with sex workers, the outreach strategy was subject to several refinements to try to ensure that the approaches responded appropriately to the ways in which gender, power, stigma, and poverty interplayed to shape their experiences. The strategy evolved to include the hiring of peer educators of different ages, the creation of the drop-in centres, the introduction of pictorial materials, and the creation of composite interventions. Similarly, the condom promotion and distribution strategy and clinical service delivery models evolved: several models were implemented, adapted, and ultimately differentiated according to the changing needs, perspectives and experiences of the clients.

The most important component of the targeted interventions was the gradual inclusion and integration of the community of female sex workers in the provision of services to be more responsive to their needs. The targeted interventions started with needs assessments that led to a better understanding of the community of female sex workers, which revealed how addressing threats of violence and harassment were more important than HIV prevention. The regular involvement of the peer educators facilitated community-led interventions, eliciting interest in forming community-based organizations and generating greater community participation. This contributed to a social movement recognizing the rights of sex workers and their social entitlements, and hence the final interventions were broader than originally planned in order to address the broader gender inequitable relationships shaping sex workers' experiences.

Data for change from a gender perspective: The targeted interventions benefited from a broad variety of data sources that were triangulated to provide information and inform implementation. The program used three major sources of data: periodic surveys and assessments; annual sentinel surveillance; and routine program data, all of which were sex-disaggregated to assess changes through time. A key lesson, however, was for managers to recognize that "the search for perfect data never ends" and that they must make "decisions based on the best available data rather than wait for the next sample or a more refined analysis", as well as be responsive to the needs of target groups (Rau 2011). A study by Kumar et al. (2011) found that a statistically significant steep decline occurred in HIV prevalence among young pregnant women in the districts with a high intensity of targeted interventions, suggesting that the interventions played an important role in bringing about the decline.

With thanks to Sameh El-Saharty and adapted from El-Saharty & Nagaraj (2015); Rau (2011); Kumar et al. (2011).

Incorporating gender analysis into implementation research outcomes
Incorporating gender analysis into implementation research outcomes considers who is empowered and disempowered by the research process and results, including the extent to which the research process itself progressively transforms gendered power relations, or at least does not exacerbate them (Morgan et al. 2016). This includes the recognition that research activities and recommendations can either aggravate or disrupt power relations, which can lead to gender and health inequities among men and women, or progressively challenge or change power relations. Health systems research which aims to progressively transform gender relations is specifically developed to consider and address inequality generated by unequal norms, roles and relations as a result of gender and other social

stratifiers. Such research incorporates aims, objectives, and/or questions that explicitly address gender and gender relations. While implementation research has a more focused aim - to explore how and why an intervention fails or works in real-world settings and improve its implementation - it still has the capacity to influence gender relations in both positive and negative ways through how it considers and incorporates gender relations into the research process and dissemination. It is important to recognize, however, that in some instances attempting to transform or influence gender relations through implementation research may be difficult due to resistance by researchers, implementers, and/or policymakers. Individuals who benefit from current power structures, for example, may try to actively avoid this type of analysis or belittle it. In these instances, implementation researchers need to carefully reflect upon how their research considers and responds to unequal gender power relations.

Implementation researchers can use the gender integration continuum to consider how their research and dissemination responds to and addresses gendered power relations (Caro 2009). The continuum categorizes approaches by how they address gender norms and relations - approaches can be either gender blind (i.e. fail to consider gender) or gender aware (i.e. consider and/or incorporate gender) (Caro 2009; Kraft et al. 2014). Gender aware approaches can be either gender exploitative, by taking advantage of "*rigid gender norms and existing imbalances in power to achieve […] program objectives*"; gender accommodating, by acknowledging "*the role of gender norms and inequities and seek[ing] to develop actions that adjust to and often compensate for them*"; or gender transformative, by actively striving "*to examine, question, and change rigid gender norms and imbalance of power as a means of reaching health as well as gender equity objectives*" (Caro 2009:10). Gender transformative approaches are most likely to address and change the underlying dynamics and structures which perpetuate inequities. Such approaches "*encourage critical awareness among men and women of gender roles and norms; promote the position of women; challenge the distribution of resources and allocation of duties between men and women; and/or address the power relationships between women and others in the community, such as service providers and traditional leaders*" (Caro 2009:10).

Take the example of an intervention which seeks to increase women's access to maternal health services. Through implementation research, researchers may find that "*gender roles and household decision-making authority [influences] the extent to which individuals access needed health services, notably where decisions relating to health seeking are deferred to the male heads of households*" (Peters, Tran, et al. 2013:19). In such cases, women might delay seeking care, even in the case of an emergency, if the male head of household is not home. Where "*such barriers do exist implementation researchers and programme managers can play an important part in changing the approach used to inform communities about the care available; for example, employing messaging about alternative decision-makers for health seeking when male heads of households are not home.*" This approach recognizes the role that gender power relations play in influencing women's access to health services, and seeks to address these by improving the implementation process. However, it does nothing to actively challenge or change the underlying problem, which is that women are unequal to men and therefore have to defer to them when it comes to (health) decision making. A transformative approach would seek to challenge the unequal gender power relations that maintain these inequities and increase women's health-seeking decision-making autonomy.

In implementation research the question or problem defines the method, and an increasing array of methodologies are being used (Peters, Tran, et al. 2013). One way in which implementation researchers can focus on transformation is through the use of participatory action research (PAR) (Loewenson et al. 2014), which is gaining traction in implementation research. PAR seeks to engage participants as active respondents who are considered to be "best placed to understand their context" and "act and reflect on self-identified problems or issues", and bring their own embedded or tacit knowledge to the analysis, development, and evaluation of the intervention (Morgan et al. 2016:8; Loewenson et al. 2014; Corbett et al.

2007). It advocates for participants' involvement throughout the research process, allowing for the potential to use the process itself to change unequal gender relations by having participants reflect upon, challenge, and alter unequal gender norms, roles, and relations. Feminist PAR, in particular, explores how unequal gender relations, and 'the centrality of male power,' leads to inequities among different groups, specifically seeking to address and challenge these within the research process (Corbett et al. 2007). Implementation researchers that use PAR can engage participants in activities that allow them to reflect upon how unequal gender relations affect implementation, and how these relations can be challenged and changed in the process of improving implementation.

Example: Participatory Learning and Action to address gendered dynamics in sexual and reproductive health and rights: Implementation research in South Sudan

The implementation problem set in context: At 2,054/100,000 South Sudan has the highest maternal mortality in the world. At marriage, women's families are paid a bride price, which in Northern Bahr el Ghazal (NBeG) is paid in cows. Girls are seen as an investment and looked after well until they are married off - often in early adolescence. South Sudan is in transition - with years of conflict and the construction of a new nation, in some areas, existing gender norms which expect women to bear many children are intensified in order 'to replace the ones that were lost'. Gender and societal norms are also in transition. Elders complain that the young no longer listen to and respect their elders, not all couples keep to the traditional three years birth spacing, and marry younger than before. Some parents worry that their girl may get pregnant while unmarried; as a result, they marry her off early, especially if she is not in school. Girls and boys have very limited access to sex education or contraceptives. These factors can lead to early pregnancy.

The implementation research approach: The South Sudan Health Action and Research Project (SHARP) (KIT undated) aims to improve maternal health and is funded by the Dutch Government and implemented by the Royal Tropical Institute (KIT), Healthnet TPO, International Medical Corps (IMC) and Cordaid, in close collaboration with the Ministry of Health. Implementation research is a core part of the approach to addressing maternal health, and in the area of community engagement (an approach building on Participatory Learning and Action), community dialogues to discuss and explore opportunities for changing gender norms were carried out.

Using community dialogue to enable reflection: KIT compiled (from various sources but drawing substantially on the GTZ developed generational dialogue (von Roenne 2012) and designed a curriculum for the training of community facilitators to support dialogue and reflection on norms and values shaping maternal mortality and health service access and use. The curriculum was further adapted with input of REACH Trust, Malawi, and SHARP partners. The training of community facilitators focussed on dialogue between older and younger women and men, comparing social and gender norms and practices between present and past, sharing knowledge on maternal health and discussing what needs to be changed for maternal health to improve.

Community dialogues and commitments to change: Through the facilitated discussions between genders and generations, statements for change were negotiated and agreed. Two examples include:

"We don't want daughters to marry or be pregnant before 18 years old and we in our family will do all we can - we want to pledge this to our family and community".

"Married women should be allowed to use contraception and have 3 year birth spacing" *(which is an incredible shift given the initial resistance by men).*

This is an inspiring example of how gender aware approaches inspired by Participatory Learning and Action can bring gender transformative change at the community level. In 2016, 36 communities participated in these activities and the approach will be further rolled out and hopefully continue to challenge and change views and practices that undermine women's maternal health.

With thanks to Kingsley Chikaphupha (REACH Trust, Malawi), Lot Nyirenda (LSTM, based in Malawi), Korrie de Koning, Egbert Sondorp & Maryse Kok (both KIT, Amsterdam). Source: (Theobald 2014)

While not all implementation research may be able to progressively transform gender relations, it is important that it does not unintentionally exacerbate them. During data collection, for example, researchers need to think about how someone's participation within the intervention or research might affect their relationships with others, such as their partners, co-workers, or community members. For example, by not considering how participation in a study can affect gender relations, an unintended consequence of some mHealth interventions was increased domestic violence, abuse, or partner control as such interventions improved women's access to information and resources, without considering men's control over these elements (Deshmukh & Mechael 2013; Jennings & Gagliardi 2013). Implementation research cannot be used only to identify such unintended consequences, it can also be used to amend the intervention and its implementation in such a way as to ensure that these consequences do not occur and/or transform the inequitable gender relations which led to these consequences. Within the above example of mHealth, implementation researchers and programme managers could, for example, have included men in implementation activities, challenging them in relation to women's unequal access to information and resources, and lack of decision-making power in regards to their use.

3. Conclusion

Incorporating gender analysis into implementation research is about analysing how gendered power relations influence the implementation of an intervention, in addition to understanding how the implementation of an intervention affects gendered power relations. Within this chapter we have outlined how gender analysis can be incorporated into health systems implementation research *content* (i.e. what gendered power relations affect effective implementation and how they affect it)*, process* (i.e. how gendered power relations affects data collection and analysis), *and outcomes* (i.e. who is empowered and disempowered by the research process and results, including the extent to which the research process itself progressively transforms gendered power relations, or at least does not exacerbate them).

By incorporating gender analysis into implementation research, researchers can ensure that gendered power relations do not prevent the successful uptake and implementation of interventions, and that implementation itself does not perpetuate existing gender inequalities. While not all intervention research needs to incorporate a gender transformative approach into its design, at the very least it should aim to ensure that unequal gender relations are not exacerbated. Recognizing gender-based constraints and power relations, as well as implicit biases within our own understanding of the world, in addition to involving both men and women within the research process from the outset, can prevent unintended consequences that are hidden by gender blind research. The inclusion of gender analysis into implementation research is therefore important if implementation is to lead to strong, equitable, and sustainable health systems and health systems interventions.

References

Bhattacharyya, O., Reeves, S. & Zwarenstein, M. (2009). What Is Implementation Research?: Rationale, Concepts, and Practices. Research on Social Work Practice. 19(5):491-502.
http://www.ibrarian.net/navon/paper/What_is_implementation_research__Rationale__conce.pdf?paperid=13616889

Caro, D. (2009). *A Manual for Integrating Gender Into Reproductive Health and HIV Programs*. http://www.igwg.org/igwg_media/manualintegrgendr09_eng.pdf .

Chilundo, B.G. et al. (2015). Relaunch of the official community health worker programme in Mozambique: is there a sustainable basis for iCCM policy? *Health policy and planning* 30 Supplement 2:ii54-ii64.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4625760/

Corbett, A.M., Francis, K. & Chapman, Y. (2007). Feminist-informed participatory action research: a methodology of choice for examining critical nursing issues. *International journal of nursing practice.* 13(2):81-8.
https://www.researchgate.net/publication/51377909_Feminist-informed_participatory_action_research_A_methodology_of_choice_for_examining_critical_nursing_issues

Datiko, D.G. et al. (2015). Exploring providers' perspectives of a community based TB approach in Southern Ethiopia: implication for community based approaches. *BMC Health Services Research* 15(1):501.
http://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-015-1149-9

Deshmukh, M. & Mechael, P. (2013). *Addressing Gender and Women's Empowerment in mHealth for MNCH: An analytical Framework*. MHealth Alliance, UN Foundation.
http://www.villagereach.org/wp-content/uploads/2013/07/gender_analytical_framework_report.pdf

El-Saharty, S. & Nagaraj, A. (2015). *Reducing the Risk of HIV/AIDS among Female Sex Workers in India. A Case Study*. Washington, DC: The Global Delivery Initiative. The World Bank.

George, A. (2008). Nurses, community health workers, and home carers: gendered human resources compensating for skewed health systems. *Global Public Health* 3 Supplement 1:75-89.
http://www.tandfonline.com/doi/abs/10.1080/17441690801892240?journalCode=rgph20

Hunt, J., 2004. Introduction to gender analysis concepts and steps. *Development Bulletin* 64:100-106.
https://ccep.crawford.anu.edu.au/rmap/devnet/devnet/gen/gen_mainstream.pdf#page=139

Jennings, L. & Gagliardi, L. (2013). Influence of mHealth interventions on gender relations in developing countries: A systematic literature review. *International Journal for Equity in Health* 12(85). https://equityhealthj.biomedcentral.com/articles/10.1186/1475-9276-12-85

Jewkes, R.K., Levin, J.B. & Penn-Kekana, L.A. (2003). Gender inequalities, intimate partner violence and HIV preventive practices: findings of a South African cross-sectional study. *Social Science & Medicine* 56(1):125-134. http://www.ncbi.nlm.nih.gov/pubmed/12435556

JHPIEGO, 2016. Gender Analysis Toolkit for Health Systems, Available at: http://reprolineplus.org/system/files/resources/Gender-Analysis-Toolkit-for-Health-Systems.pdf [Accessed April 19, 2016].

KIT (undated). South Sudan Health Action and Research Project (SHARP). KIT Health. http://www.kit.nl/health/project/south-sudan-health-action-research-project-sharp/

Kraft, J.M. et al. (2014). An evidence review of gender-integrated interventions in reproductive and maternal-child health. *Journal of health communication* 19 Supplement 1:122-41.

Kumar, R. et al. (2011). Impact of targeted interventions on heterosexual transmission of HIV in India. *BMC Public Health* 11(1):549. http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-11-549

Larson, E. et al., 2016. 10 Best resources on... intersectionality with an emphasis on low- and middle-income countries. Health Policy Planning Advance Access doi:10.1093/heapol/czw020. http://heapol.oxfordjournals.org/content/early/2016/04/27/heapol.czw020.abstract

Loewenson, R. et al. (2014). Participatory Action Research in Health Systems: A Methods Reader. Harare: TARSC, AHPSR, WHO, IDRC Canada, EQUINET. http://www.equinetafrica.org/sites/default/files/uploads/documents/PAR%20Methods%20Reader2014%20for%20web.pdf

LSTM, 1996. Guidelines for the analysis of Gender and Health. Liverpool School of Tropical Medicine. https://www.lstmed.ac.uk/sites/default/files/pictures/Guidelines%20for%20the%20Analysis%20of%20Gender%20and%20Health.pdf .

Morgan, R. et al. (2016). How to do (or not to do)… gender analysis in health systems research. Health Policy and Planning Advance Access doi:10.1093/heapol/czw037. http://heapol.oxfordjournals.org/content/early/2016/04/26/heapol.czw037.abstract

Nowatzki, N. & Grant, K.R. (2011). Sex is not enough: the need for gender-based analysis in health research. Health Care for Women International, 32(4):263-277. https://www.researchgate.net/publication/50401590_Sex_Is_Not_Enough_The_Need_for_Gender-Based_Analysis_in_Health_Research

Östlin, P. et al. (2006). Gender and health promotion: a multi-sectoral policy approach. Health promotion international 21(suppl_1):25-35. http://heapro.oxfordjournals.org/content/21/suppl_1/25.short

PAHO (2009). *Guide for analysis and monitoring of gender equity in health policies*. Washington, DC. Available at: http://new.paho.org/hq/dmdocuments/2009/Guide_Gender_equity_.pdf

Peters, D.H., Adam, T., et al. (2013). Implementation research: what it is and how to do it. *BMJ* (Clinical research ed.) 347:f6753. http://www.bmj.com/content/347/bmj.f6753

Peters, D.H., Tran, N.T. & Adam, T. (2013). *Implementation Research in Health: A Practical Guide*. World Health Organization: Geneva. http://www.who.int/alliance-hpsr/resources/implementationresearchguide/en/

Rau, B. (2011). *The Avahan-India AIDS Initiative: Promising Approaches to Combination HIV Prevention Programming in Concentrated Epidemics*. Arlington, VA: USAID's AIDS Support and Technical Assistance Resources (AIDSTAR-One) Task Order 1. https://aidsfree.usaid.gov/sites/default/files/avahan_sw_india.pdf

Ravindran, T.K.S. & Kelkar-Khambete, A., 2007. Women's health policies and programmes and gender-mainstreaming in health policies, programmes and within health sector institutions, Available at: http://www.who.int/social_determinants/resources/womens_health_policies_wgkn_2007.pdf.

RinGs (2015). *Ten Gender Analysis Frameworks & Tools to Aid with Health Systems Research*. http://resyst.lshtm.ac.uk/resources/ten-gender-analysis-frameworks-tools-aid-health-systems-research .

Von Roenne (2012). How to organise Generation Dialogues about Female Genital Cutting. GTZ. https://www.giz.de/expertise/downloads/Fachexpertise/giz2013-en-fgm-GenDia-man-t1.pdf

Sanders, D. & Haines, A., 2006. Implementation research is needed to achieve international health goals. *PLoS medicine*. 3(6):e186. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1472551/

Sen, G. et al. (2007). *Unequal, Unfair, Ineffective and Inefficient Gender Inequity in Health: Why it exists and how we can change it*. World Health Organization, WHO Commission on Social Determinants of Health. http://www.who.int/social_determinants/resources/csdh_media/wgekn_final_report_07.pdf .

Thapa, D.K. & Niehof, A. (2013). Women's autonomy and husbands' involvement in maternal health care in Nepal. *Social Science & Medicine* 93:1-10. https://www.researchgate.net/profile/Deependra_Thapa/publication/254263385_Women's_autonomy_and_husbands'_involvement_in_maternal_health_care_in_Nepal/links/00b7d520ec9452a2dc000000.pdf

Theobald, S. (2014). *Community engagement to tackle maternal mortality in South Sudan*. ReBUILD. https://rebuildconsortium.com/blog-news/blogs/2014/community-engagement-to-tackle-maternal-mortality-in-south-sudan/

Theobald, S., Simwaka, B.N. & Klugman, B., 2006. Gender, health and development III: engendering health research. Progress in Development Studies, 6(4), pp.337-342.

Thys, S. et al. (2015). Why latrines are not used: communities' perceptions and practices regarding latrines in a Taenia solium endemic rural area in Eastern Zambia. PLoS neglected tropical diseases 9(3):e0003570. http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003570

Tulloch, O. et al. (2015). Patient and community experiences of tuberculosis diagnosis and care within a community-based intervention in Ethiopia: a qualitative study. BMC Public Health, 15(1):187. http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-1523-x

WHO (2016). What do we mean by 'sex' and 'gender'? Available at:
http://apps.who.int/gender/whatisgender/en/ [Accessed March 28, 2016].

Yassin, M.A. et al. (2013). Innovative community-based approaches doubled tuberculosis case notification and improve treatment outcome in Southern Ethiopia. PloS one 8(5):e63174.
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0063174

Chapter 12: Achieving influence and driving change

Fatema Rajabali and Hannah Corbett
Institute of Development Studies

"*It is clear that better use of research in development policy and practice can help save lives, reduce poverty and improve quality of life*". (Court and Young 2003: p1)

"*There is nothing a government hates more than to be well informed; for it makes the process of arriving at decisions much more complicated and difficult*" (Maynard Keynes Collected Writings, vol. 21, p. 409).

"*Policy-makers, managers, and funding agencies do not always want to know how their programmes are being implemented, unless of course they can be shown to be doing well. They may have invested considerable political and financial capital in a policy, and be afraid of not producing the desired results or of poorly managing resources. Funders are frequently resistant to research that might highlight sustainability issues or the negative unintended consequences of their programmes, such as the human resource distribution problems arising as a result of hiring people for single purpose projects.... Similarly, the concerns of minority groups may not be of interest to those groups in power, particularly if there are social and political sensitivities. … In some cases participatory action research may even be considered revolutionary to the existing power structures. In these circumstances, an important aspect of implementation researchers' work is to find ways to get their research into agenda-setting processes to influence policy. This may also require approaches that rely more on advocacy strategies that can make use of well-designed research.*" (Peters et al. 2013: p42).

1. Research knowledge driving change

There has been growing awareness and acknowledgement of the value of research uptake and communications as an integral part of the research process. Bringing together good credible research is important, but there is increasing demand to see it come alive rather than just have it sit on a database somewhere. The World Health Organization (WHO) has also recognised the need to use rigorous processes to ensure that health recommendations are informed by the best available research evidence (Al-Riyami 2010).

In order to see relevant research make a difference requires putting the evidence into use through a variety of knowledge management and communication mechanisms – an area this chapter will explore. A more detailed focus will be presented on the use of policy briefs, an evidence-based product that places strong emphasis on clear recommendations for policy-related professionals. But apart from just communicating credible evidence, researchers also need to be more engaged and aware of the complexity of decision-making processes, as this directly influences the likelihood of whether their research will be taken up. This is another area that we will briefly examine in this chapter.

There are a number of terms that we will be using that are explained below.

Knowledge: is a well-researched concept and refers to information that is available and accessible. Perkin and Court (2005) define knowledge as "information that has been evaluated and organised so that it can be used purposefully" (p2). Porter (2010) reflects that the term is now often used in place of evidence as it encourages discussion on how evidence is processed and utilised, and includes tacit and informal sources.

Evidence: in the context that we are discussing, evidence is information generated through research, whether scientific or social, and generally communicated through research-related formats including data, statistics, indicators, scientific studies, technical briefings and reviews. In this chapter, we will focus on research evidence and the process by which it can be communicated and utilised by decision-makers including the policy community.

Evidence-informed policy: can be considered as "*that which has considered: a broad range of research evidence; evidence from citizens and other stakeholders; and evidence from practice and policy implementation as part of a process that considers other factors such as political realities and current public debates*" (Newman et al. 2012:p17). Evidence-informed policy does not generally involve a linear transition of research findings into policy decisions; research can inform policy discourses in multiple and sometimes subtle ways. But it is valuable to understand the complex nature of decision-making processes and where your research may fit in. Even when it does not directly contribute to decision-making, it may influence the language used by policy-makers or create awareness of an issue (Weyrauch and Diaz Langou 2011). Effectively used, research evidence can have a considerable impact:

> "*… initial outcomes of the study which wound up at the end of 2011 showed increased community awareness about benefits of delivering in health facilities, and phenomenal increases in facility births, with an average of 1,336 deliveries per month in the intervention area compared to an average of 461 deliveries per month in the control area*". (FHS 2014)

2.   Literature around the research - policy nexus

The evidence-based policy concept has been well explored in the international development sector with multiple sources of literature and the development of useful models and frameworks that have helped to improve research – policy integration and research uptake by policy actors. Annex table 1 outlines some selected bodies of work that address the use of research evidence in policy.

Although important strides have been taken in raising expectations and pushing boundaries, progress is not uniform. A recent UK/Dutch workshop in London (Wellcome Trust 2014) identified that while some research organisations have led the way in exploring research uptake and knowledge management by strengthening capacity, creating alliances and pioneering new ways of working, many others still struggle. There is still much to be done in mastering the 'art' of research uptake and making sure it happens more consistently (Morton 2015).

3.   Do we understand enough about real world policy processes?

"*There is no such thing as context-free evidence*" (Davies 1999:p111).

Politics shapes how evidence is used at many decision-making levels - and this means that researchers need to understand politics and the process of decision-making. Power relations can crowd out certain types of evidence and perspectives (Fisher and Vogel 2008), so engaging with political actors and understanding how decisions are made is essential. Porter (2010) provides some useful guidelines for integrating political economy analysis into different stages of the research and communication process in order to negotiate the political context.

Achieving and attributing influence and change is a complex and difficult process with no easy quick-wins. As Dagenais (2015) argues "*despite efforts expended over recent decades, there is a persistent gap between the production of scientific evidence and its use*" (p1654). It is also

important to take account of how definitions of 'successful' influence and impact are shaped and by whom. Scott-Villiers highlights this in her case study of research undertaken in the Karamoja, Uganda which was deemed as 'not influential' by a donor, but had a significant effect on the local community (Scott-Villiers 2012). Debates around influence and impact often focus narrowly on 'policy influence' and 'policymakers' (in itself a very broad term that encompasses a whole range of different actors and that requires further definition in relation to specific contexts). However, as Benequista (2016) argues in a recent blog on research communications and politics, policymakers are not the only drivers of change; in fact they can obstruct positive change. An analysis of the power dynamics and relations in the context in which researchers are seeking to engage and achieve influence is essential (Gaventa 2006) in order to realise change and avoid simply reinforcing or legitimising existing power imbalances and the status quo.

This is further analysed by Lomas (2006), who asks how much researchers should compromise in their conception of 'evidence' and how much should decision-makers compromise in theirs? In some cases it is possible for researchers and decision-makers to be equal partners in the co-production of research (Lomas 2005), rather than the latter being seen as the end-users of findings generated by researchers (Pope et al. 2005). Porter (2010) suggests that practice to date shows that involving governments in the process of establishing research priorities increases the likelihood of the uptake of findings by those governments. It also makes it more likely that recommendations are not out of line with government thinking and are mindful of the realities of policymaking.
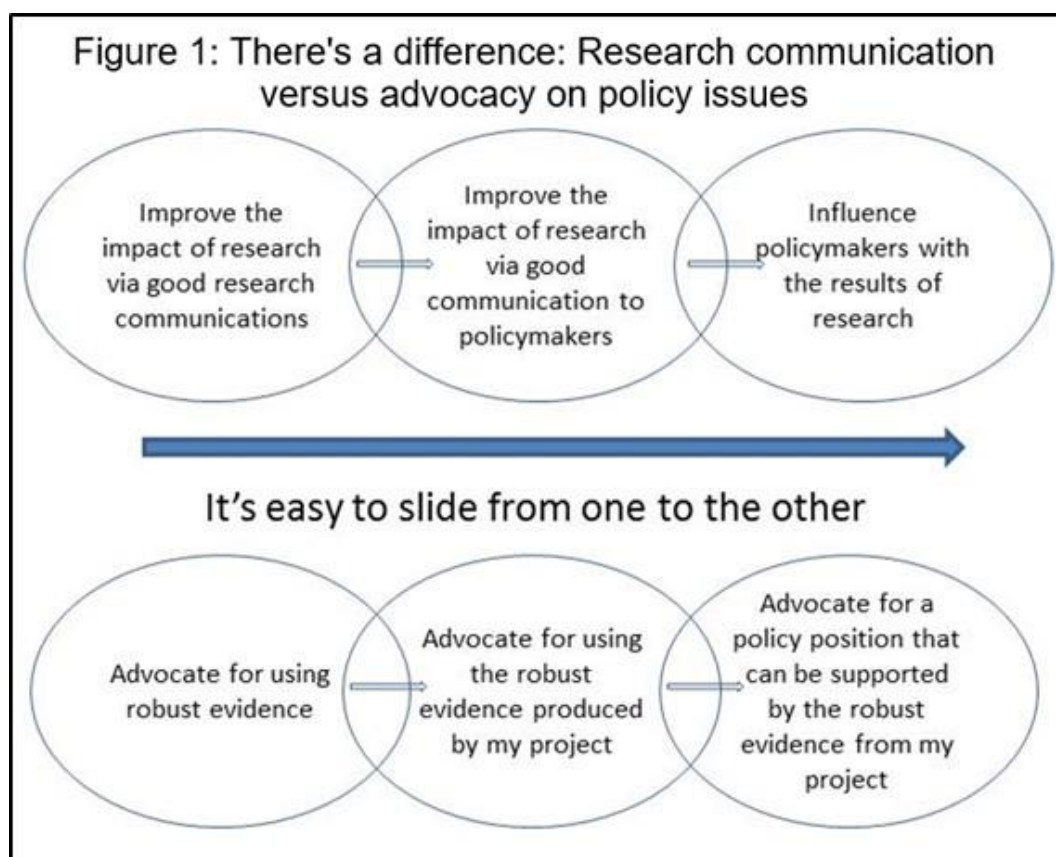
Roger (2006) discusses the need to confront the gap between the idealised use of research in policy development and current realities. He highlights that healthcare managers and decision-makers do not function solely within the simple paradigm of 'What works?'. Their questions more often take the form: 'What combination of interventions works where, for which sub-populations, in which environmental circumstances, in which combinations?' As Porter (2010) argues, policymakers often want research that shows how impacts can be achieved. They are seeking evidence that demonstrates how things should be done differently or that offer practical guidance. How policymakers define 'useful' research will often depend on whether the evidence helps them solve a policy problem. This suggests that decision-makers and researchers need to negotiate ways of meeting halfway in this process, which Greenhalgh and Russell (2005) describe as "*a new rationality of policy-making*" (p40).

But there is also a need to go beyond just engaging with policy actors themselves. Dutta (2012) argues that researchers are knowledge producers and communicators, and "*if they view their role to policy, they should be prepared to engage with stakeholders affected by policy issues and expose their findings to human interaction, review and scrutiny by others*" (p9).

4.   Your research needs to have impact! What can this mean?

Realising this kind of positive transformative change is for many researchers and research institutions at the heart of what they do. In recent years, the demand on the research community to ensure the impact of research findings on the decisions, actions and behaviours of policymakers and practitioners has become an increasing priority. This has been driven in part by growing funder requirements and expectations (Sumner et al. 2009), and contested value for money agendas (Chambers 2014). Shaxson (2012) discusses how researchers are being put under increasing pressure to demonstrate impact and examines what that impact could look like, highlighting four very useful points:

1. Clear research evidence does not necessarily lead to clear policy messages. It can be better to focus on the quality of the evidence produced by getting it into debates, rather than trying to demonstrate impact in terms of any concrete outcomes on policy.
2. Be careful how you define 'policy relevance'. Although the term is frequently used, it has can have multiple dimensions depending on the type of research being conducted. Often, as Shaxson notes, policy relevance is not an either/or situation; it's a multidimensional and constantly shifting challenge.
3. Be realistic about what can be achieved - think breadth of impact rather than depth. Yes, research can influence policy development, but it is important to recognise that it can also affect the process by which policies are designed and implemented, and the relationships between researchers and other stakeholders. Pankhurst (2012) goes further by arguing that it is probably less important to ensure that research leads to policy change than to support local systems for reviewing and evaluating evidence, which may allow decision- makers to be better able to set that evidence in the context of what needs to be done.
4. Be clear whether you are practising research communication or advocacy. As shown in the diagram below (Shaxton (2012) there should be a clear line between these two activities. Researchers and organisations need to decide where they sit, especially on which side of the line and how far to the right of the diagram they wish to operate, and to be very open about this from the outset.

Figure 1: There's a difference: Research communication versus advocacy on policy issues

Improve the impact of research via good research communications

Improve the impact of research via good communication to policymakers

Influence policymakers with the results of research

It's easy to slide from one to the other

Advocate for using robust evidence

Advocate for using the robust evidence produced by my project

Advocate for a policy position that can be supported by the robust evidence from my project

5.   I have completed my research: So, now what?

As presented earlier, the generation of credible research is in itself insufficient if you are looking to influence change. Your research methods and results need to be intelligible to non-researchers, sufficiently digestible, and have results clearly interpreted and translatable for the target audience. Many decision-makers, especially policy-related professionals, often

have little time for engaging with research and if the results are ambiguous and do not offer clear interpretation of the findings, they will be discarded, ignored or misinterpreted (Porter 2010). The produced research usually needs to be packaged in a variety of products to make it more accessible. One can use working papers, briefing papers, policy briefs, talking head videos, etc. as a way of making a much larger, more detailed research report more useful for the audience at hand.

Using policy briefs to communicate your research
Built of the assumption that policy informed by evidence is more likely to lead to better development outcomes (Newman et al. 2012), we would like to explore the use of the policy brief as one research communication approach. A policy brief is a concise summary of a particular issue, the policy options to deal with it, and some recommendations on the best option. It is aimed at government policymakers and others who are interested in formulating or influencing policy as they take decisions in complex policy processes (Beynon et al. 2012).

There are a number of points to consider when you are developing policy briefs:

1. **Who is your target audience?** Are they policy-related advisors? If not, do you need to develop a policy brief or is another knowledge product more suitable?
2. **What are your intended policy impacts?** What kind of change are you hoping to see as a result of your policy brief? Are you simply looking to create awareness on a policy issue, influence decisions around programming and funding, or looking for a change in behaviour?
3. **The audience may not be health and development specialists**. The language needs to be well-thought through without overly simplifying your key messages. The use of jargon and technical terminology needs to be avoided if possible or at least clearly explained.

Table 1: Policy brief structure

Executive statement	Designed to give an overview of the content of the brief with an emphasis on capturing the attention of the reader.
Introduction	Explains the importance and urgency of the issue and creates curiosity about the rest of the brief. It gives an outline with the structure for the brief and an overview of the conclusions or the direction of the rest of the brief.

Methodology	Aims to strengthen the credibility of the brief by explaining how the findings and recommendations were arrived at, including:
	• Description of the issue and context of the investigation;
	• Description of the research and analysis activities;
	  - What methods were used to conduct the study?
	  - Who undertook the data collection and analysis?

Results and conclusions	1. Provides an overview of the findings/facts constructed around the line(s) of argument behind the policy recommendations moving from general and specific information. Base the conclusions on evidence, data and findings with clear, balanced and defensible assertions.
Implications	2. Attempts to explore what policy changes or actions the results point to, based on the evidence provided. It is less direct that recommendations and useful to include if direct advice is not requested or welcome.

Recommendations	Highlights what you as a researcher think should happen, based on the evidence presented. It is useful to make the recommendations actionable, with some precise steps of what should happen next, so the key target audience have some guidance on steps forward.
References and useful resources	3. It is important to cite and clearly signpost the sources in your briefing to provide credibility in your analysis and recommendations. This gives authority and weight to your product. Readers generally find additional resources useful to have if they would like to undertake more detailed reading around the subject area.

Source: Knezovich (2014)

Additional points to consider:

1. Policy briefs are two, four or a maximum of eight pages (1,200, 2,200 or 4,000 words);
2. Policymakers often spend around 30–60 minutes reading information on an issue;
3. Design can help highlight key facts or concepts;
4. The policy brief should be focused, succinct and make *appropriate* recommendations;
5. In addition to having solid content, policy briefs should also be *visually engaging*;
6. Accessibility is important - the style should be professional - not academic. Remember your target audience. Accessibility includes making sure that technical language is well explained.

6.   What tools can one use in achieving influence and driving change?

There are a range of research uptake and communication tools and approaches that can be used to better support you in getting your evidence and your knowledge products (such as a policy brief) out in the health and development practice and policy world. As noted later in the chapter, any approach used should be developed as part of a wider action plan. There are different activities that a researcher can undertake with the research evidence that has been produced:

1. Sharing information: primarily one-directional - less need to capture feedback;
2. Engagement: when you might be looking for feedback or to stimulate conversation on an issue;
3. Collaboration: seeking to work in with other actors towards a shared objective;
4. Storing/capturing information: in ways that facilitate easy future access.

Multiple tools and a mixture of facilitation techniques can help to foster effective knowledge-sharing of research evidence. Annex table 2 outlines a range of tools, with examples from the health sector, that can be used by researchers in developing a research uptake strategy.

Example: Briefing papers to communicate research findings

The aim of the *Joto Afrika* briefing series, co-produced between multiple organisations, was to help people (particularly decision-makers) understand the issues, constraints and opportunities that poor people face in adapting to climate change and escaping poverty, by translating information in a way that would meet the demands of stakeholder groups, including communities, researchers, and policymakers.

In an evaluation survey over 90 per cent of 319 respondents noted experiencing personal learning while 45 per cent indicated taking action in their own work (beyond just learning) as a result of reading *Joto Afrika*. Many provided examples of that action. During the UN Climate Change Conference in Copenhagen (COP15), the then Kenyan Minister of Water and Irrigation, Dr Charity Ngilu, used *Joto* Issue 2 on water issues during her speech and pointed out a case study from her home area that featured a water harvesting project. Since then she has implemented various water projects in several parts of Kenya.

Example: use audio-visual tools to bring your research to life

Digital storytelling can capture 'stories of change' that have emerged either directly or indirectly from a research process. Building on the briefing series described above, a photo-audio story was produced. This is a form of digital storytelling where photographs are presented alongside a narrative (usually in the first person) giving communities a voice to tell their story.
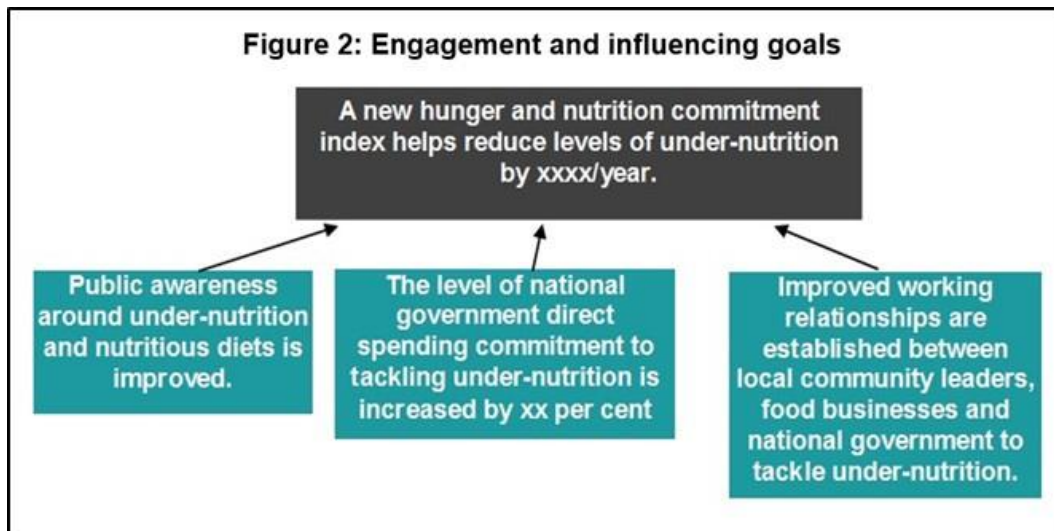
The use of talking heads is another visually stimulating and engaging approach to getting an expert to explain why the research or key issue is important. In this example, Dr David Heymann discusses new disease threats and the most effective ways to address them

Audio podcasts can also be a useful tool to share key messages about the research. This podcast from the London International Development Centre (LIDC) focuses on the development of assessment tools for impact evaluation of malaria control for children in Mali and Senegal.

7.   Setting engagement and influencing goals and objectives

Despite the complexity, effective action is possible (Jones et al. 2012) if researchers can involve both potential beneficiaries and other actors including project or programme managers, government officials, health practitioners and community leaders (Douthwaite et al. 2008; Douthwaite 2007).This section of the chapter will outline practical approaches and tools and tactics that can help researchers design effective influencing and engagement strategies in relation to their work, with a particular focus on how to identify, understand, prioritise and target key audiences.

Defining a clear and measurable engagement and influencing goal, underscored by some narrower objectives, is a critical first stage of the design of any influencing and engagement plan. These objectives need to be consistent and flow from the intervention *theory of change*, (Vogel 2012), as discussed in chapter 4. The goals and objectives could relate to an institutional or organisational priority, a research programme or project or a specific output such as a research report. They will help inform decisions around which particular organisations or individuals such as activists, NGOs or government officials you seek to engage with and influence, and how, in order to achieve the change you want to see.

Figure 2: Engagement and influencing goals

A new hunger and nutrition commitment index helps reduce levels of under-nutrition by xxxx/year.

Public awareness around under-nutrition and nutritious diets is improved.

The level of national government direct spending commitment to tackling under-nutrition is increased by xx per cent

Improved working relationships are established between local community leaders, food businesses and national government to tackle under-nutrition.

8.   Understanding your audiences and their operating environments

If researchers aim to engage in dialogue through structured processes, experience has shown that careful planning is required to clarify intentions, select who to engage with, when to engage, and how best to do so (Dutta 2012:9-15). Once you have your engagement and influencing goal and objectives, you can start to think about which individuals and organisations will be critical in helping you achieve them. Thinking beyond individuals and organisations in isolation, and taking into account the broader international and national political, social, economic and cultural environments they are operating in and the power dynamics and the relationships between you and your audiences and their relationships with each other is critical (Court et al. 2004). Demonstrating this understanding in your engagement and influencing tools and outputs will be critical to establishing credibility amongst your key audiences, and ensuring the uptake of your research findings.

The stakeholder mapping process
There are various stakeholder mapping tools that can be used to help researchers identify and prioritise the audiences that they will need to engage with and influence in order to achieve their policy-influencing and engagement goals and objectives. The following diagram outlines the general stages of the mapping process that need to be undertaken.

Figure 3: Stakeholder mapping process
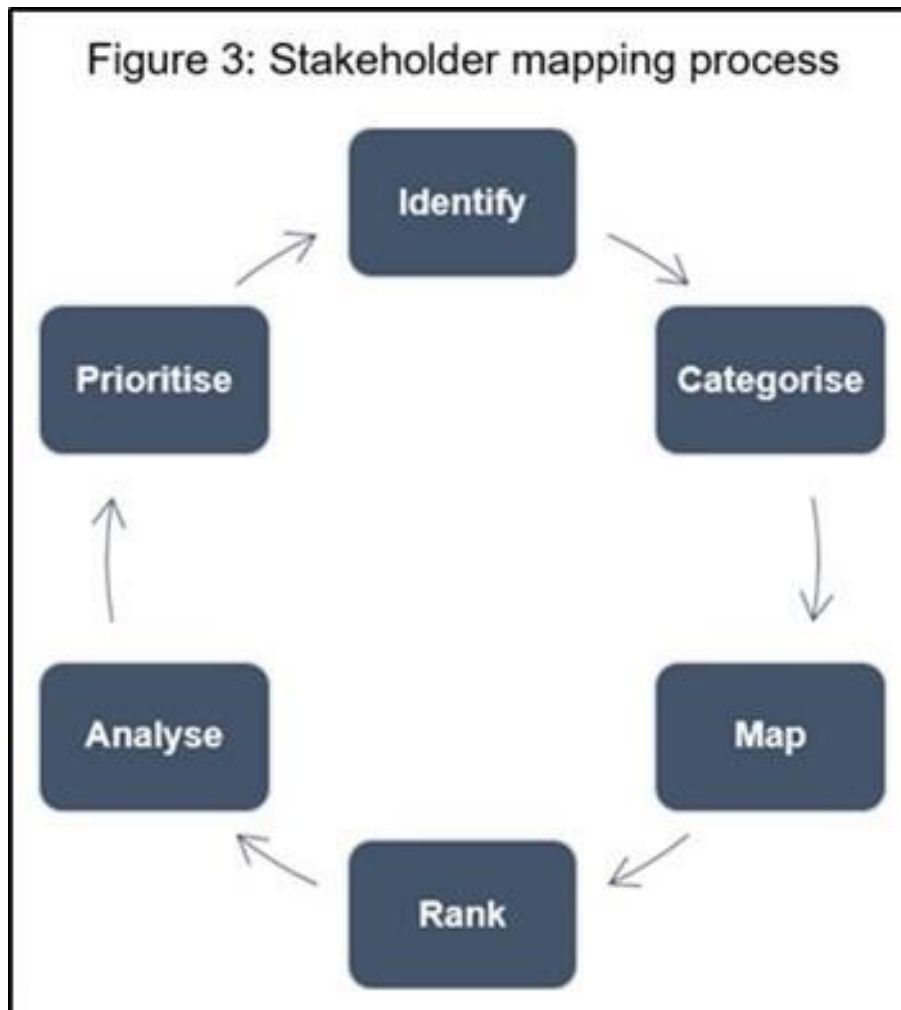
In relation to your influencing goals and objectives:

- **Identify** stakeholders (individuals, organisations);
- **Categorise** stakeholders by type (government, media, donors);
- **Map** relationships and links between stakeholders;
- **Rank** stakeholders (by influence, power, alignment, interest, attitude);
- **Analyse** stakeholders' positions, perspectives, links and relationships, how you might want them to change, and what this might mean for your strategies to engage audiences;
- **Prioritise** your key audiences.

As discussed in chapter 5, stakeholder mapping should be an ongoing process throughout the implementation period. The process needs to reflect the dynamic spaces where relationships and power balances are constantly shifting, and needs to be reviewed and updated on a regular basis.

Example: Participatory Impact Pathways Analysis

The Participatory Impact Pathways Analysis (STEPS Centre undated) offers researchers the opportunity to come together with partners, beneficiaries of their research and other key stakeholders, including government officials and practitioners to define and visualise how they are going to achieve their policy engagement goal. In 2013, the Institute of Development of Studies working with Practical Action, and funded as part of the International Development Research Centre's (IDRC) Think Tank

Initiative's Policy Engagement and Communications Programme hosted a workshop for South Asian think tanks. As part of this workshop the think tanks, some of whom were more research-focused and some of whom were more advocacy- and policy-orientated, worked to define their own institutional engagement and influencing goal and undertook a PIPA exercise in relation to this goal. The results, including the impact stories of pathways to change that the groups came up with, were hugely varied and offer an interesting insight into the PIPA process (Georgalakis 2014). The discussions they had as part of the process of listing, grouping and analysing their stakeholders were extremely valuable.

Stakeholder mapping tools

There are a number of matrix and network mapping tools that can be used to undertake the stakeholder mapping process. The following links provide detailed explanations of some of these tools:

- Power/Interest Matrix
- Alignment, Interest and Influence Matrix
- Participatory Impact Pathways Analysis
- Net Mapping

9.   Developing your engagement and influencing plan

As part of the stakeholder mapping process you will have already started to think about how you can reach some of your audiences in terms of their relationships to you and your organisation, and also in terms of their relationships with each other. An effective engagement and influencing strategy will need to include a number of essential questions in relation to a specific audience group including:

HOW do they access information and who or what influences them?

Different audiences access information and evidence in a variety of ways. A recent study by the University of Manchester of how UK civil servants engage with academic research and expertise (Talbot and Talbot 2014) found that they preferred research information that had been 'pre-digested' in the form of a briefing or a media report. However, they also found that just over half of the respondents were accessing more traditional academic outputs such as peer-reviewed journals.

WHO is best placed to communicate with them?

This question helps to focus thinking on the relationships and leverage that exist within your organisation in relation to your target audiences and about the capacity and resources you have to act upon these. Benyon et al. (2012) consider the effect of a policy brief that included an opinion piece from a sector-recognised expert on changing behaviours and prompting actions. The study found that including the opinion led to an increase in sharing the brief more widely, but not necessarily in changing the existing perceptions and attitudes of readers.

It is important to note that others outside your organisation, such as research partners, might be better placed to act as knowledge brokers or knowledge intermediaries to communicate with a specific audience. This may lead to questions about whether they have capacity to act in this way, irrespective of how well placed they are in terms of physical location and access. There is an interesting case study of a Knowledge Broker programme implemented in Burkina Faso to strengthen the way that scientific knowledge was made available to health practitioners and policymakers (Dagenais et al. 2015).

WHEN is the best time to engage with them?

The response to this question should be shaped by internal factors such as the research timeline programme or moments in an institutional strategy, and by debates and policy windows in the external environment. The Overseas Development Institute (ODI) ROMA guide to policy influence and policy engagement (ODI 2014) outlines some useful steps to help map the external environment in relation to national government-driven policy formulation and change. However, it is important to note that national governments and the formal policymaking process are only one part of how change happens, and you will need to look to debates and activities being led by multilateral organisations, community activists, local governments and NGOs to develop a fuller picture to guide your own timeline of engagement.

WHAT do you think the best tool or tactic will be to reach them?

A wide range of different tools and tactics will be required to target and reach different audiences, which will need to be used in combination. In trying to ensure effective capturing and sharing of learning from a UNICEF and IDS social protection research programme, researchers found that "multiple media was required" (Perkins and Batchelor 2011:7).

WHAT is the expected outcome – the change in behaviour or policy that you wish to see?

It is important to ensure that your activities remain focused and that you can measure how successful they are. It could be useful, where possible, to include some broad measures of success within these, such as percentage increases in spending, Minister of Health attends and speaks at an event, citations in media or parliamentary debates.

The following table sets out these questions and an example of how you might respond to them in relation to a particular audience that you have identified.

Table 2: Policy-influencing and engagement plan example

Audience	HOW do they access information and who or what influences them?	WHO in the organisation or programme should communicate with them?	WHEN is the best time to engage?	WHAT do you think the best tool or tactic will be to reach them?	WHAT is the expected outcome, change in behaviour or policy that you wish to see?
Special Advisor to the Minister of Health	• Policy briefings • Weekly MOH internal emails • National newspapers and particular journalists • University to which they have an affiliation • Colleagues at the Finance/Treasury Department	• Director of research organisation has an existing relationship • Research partner based at university to which they have an affiliation	• Over next six months in run up to national budget • Prior to attendance at international meeting on nutrition	• Cost analysis • Briefing note in advance of meeting • Face-to-face meeting • Invitation to speak at country launch of new hunger and nutrition commitment index	• Greater political will to tackle under-nutrition • Increase in Ministry of Health direct spending on under-nutrition

10. Assessing the success of your engagement activities

As the ODI ROMA guide mentioned above highlights, traditional monitoring and evaluation approaches "*which rely on a simple feedback model with predefined indicators, collecting data and assessing progress towards pre-set objectives - are simply not adequate in the context of policy-influencing interventions*" (Young et al. 2014:44). This is something that Benyon et al. (2012) reflect on in their study of the effectiveness of policy briefs. They highlight that the simple linear model of actors such as government officials receiving policy-relevant messages, taking action upon these messages, leading ultimately to improved lives, vary rarely plays out in real life. Changes in behaviour and attitudes are also difficult to capture, especially through quantitative data. This is where impact stories and narratives can help illustrate your reach, influence and impact.

However, practical steps towards measuring the success and impact of influencing and engagement activities are possible and could include the following:

1. **Review your stakeholder map:** have the positions of stakeholders, in terms of their relationship to the research, your organisation or to each other, changed, and can this be attributed to your policy-influencing and engagement objectives?
2. **Measure your success against your defined indicators:** how far have you been successful in some of the broad indicators of success outlined in the expected outcomes section of your engagement plan? Scott and Munslow (2015) highlight some useful approaches to tracking research and policy conversations in online spaces.
3. **Capture the results of your activities in impact stories:** attribution in relation to policy-influencing is complex and difficult. Quantitative measures are one part of this process. However, using narrative in the form of the written word or multimedia content (FHS 2012), can help bring your stories of change to life.

Annex

Table 1: Selected bodies of work on research evidence

Body of work	Overview of research questions	Examples of research actors
1. Approaches and frameworks for connecting research and policy	• What processes mediate and facilitate the use of evidence and knowledge in policymaking? • How does evidence contribute to effective policymaking?	• *Overseas Development Institute RAPID framework of research - policy linkages* (Court and Young 2003). Key influences identified: (1) Political context and institutions (2) Credibility and communication of the evidence (3) Links, influence and legitimacy (4) External influences. • *The British Government* Cabinet Office views the use of evidence as one of eight core competencies of professional policymaking (Cabinet Office 1999).
2. Assessing the impact of research and research communication	• What factors seem to matter (or not matter) for increasing the impact of development research on policy? • What role does the communication of research play?	• The *International Development Research Centre (IDRC)* has analysed different methodologies for assessing the impact of research on policy and examining the challenges of assessing impact (Card.en 2009). • The *UK Department for International Development (DFID)* commissioned working papers and developed strategies /guidance notes in the area of research communications and research uptake (DFID 2008; Yaron and Shaxson 2008).
3. Theories of policy influence and models of policy change and policy processes	• What are the processes by which policy decisions are made? • How do political processes determine decisions? • Models of the policy process contain assumptions in relation to how evidence is used in policymaking.	• *Lindquist* (2001) The nature of decision-making can vary considerably. • *Weiss* (1979) How policymakers engage with and 'use' evidence: Enlightenment model; Problem-solving model. • *Stone* (2001) Bargaining and coalition formation lead to policy formulation. • *Roe* (1991) Dominant narratives can shape problem-definition and open or close off political space.
4. Models and guidance for research utilisation	• How is research consumed by policymakers? • What are the different factors that influence how policymakers demand and utilise research?	• *Weiss* (1979) *six models that explain different types of research utilisation*: Knowledge-driven; problem-solving; interactive; enlightenment; political; tactical. • *Stone* (2002) outlines 12 perspectives for improving research utilisation. These can be summarised into three categories of explanation; supply-side, demand-led and policy currents. • *Caplan* (1979) *'Two Communities' theory of under-utilisation of research* focuses on the cultural gap between researchers and policymakers. Proposes two types of research use: instrumental and conceptual use.

Source: Adapted from Porter (2010): tables 1 and 3.

Table 2: Tools for developing a research uptake strategy			
Sharing information	Collaboration	Engagement	Storing/capturing
Face to face Email email lists: e.g. WHO email listserve Discussion lists: e.g. Healthcare for all by 2015 Online communities/communities of practice (CoP): e.g. Global Health Twitter (#health, #GlobalDev) Web gathering Webinars Moodle Blogs Websites Open data	Face to face Wikis Google docs Skype Confluence Storify Facebook Google hangouts Moodle	Face to face Skype Teleconferencing E-discussions: e.g. UN-moderated e-discussion on Global Public Health Blogs Online communities: e.g. Healthboards Twitter Webinars Learning modules: e.g. Canadian Network on Health in Development	Repositories: e.g. HighWire Press Databases: e.g. InfoSci-Medical Database Social bookmarking Websites Minutes Open data Tagging

Other useful online health-related resources

Evidence-informed Policy Network (with WHO)

Exploring the Impact of Research Communications - What Difference Does a Policy Brief Make?

Loewenson, T. (2014) *Annotated Bibliography of E-Platforms Used in Participatory and Peer Exchange and Learning*, Harare: Equinet,

Overseas Development Institute ROMA guide to policy influence and policy engagement

ResUp MeetUp Symposium and Training Exchange 2015

Research Communicators: Let's Talk Politics, Shall We?

Research Gate Research sharing website

K4Health Research Utilization Toolkit

References

Al-Riyami, A. (2010). Health Researchers and Policy Makers: A need to Strengthen Relationships. *Oman Medical Journal* 25(4):251-52. www.ncbi.nlm.nih.gov/pmc/artciles/PMC3191660/ (accessed 23 March 2015)

http://www.ids.ac.uk/publication/learning-about-how-to-communicate-learning-from-and-within-a-multi-sited-organisation

Benequista, Nicholas (2016). *Research communicators: let's talk politics, shall we?* ResUp MeetUp. http://www.resupmeetup.net/research-communicators-lets-talk-politics-shall-we/

Beynon, Penelope, Christelle Chapoy, Marie Gaarder and Edoardo Masset (2012a). *What Difference does a Policy Brief Make?* Institute of Development Studies and the International Initiative for Impact Evaluation. http://www.3ieimpact.org/en/evaluation/policy-influence/policy/

Benyon, P., Gaarder, M., Chapoy, C. and Masset, E. (2012b). Passing on the Hot Potato: Lessons from a Policy Brief Experiment. *IDS Bulletin* 43.5, Brighton: IDS. https://opendocs.ids.ac.uk/opendocs/handle/123456789/7517

Cabinet Office (1999). *Professional Policy Making for the Twenty First Century.* UK Government. http://dera.ioe.ac.uk/6320/1/profpolicymaking.pdf

Caplan, Nathan (1979). The Two-Communities Theory and Knowledge Utilization. American Behavioural Scientist 22(3):459-470. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/66698/10.1177_000276427902200308.pdf?sequence=2

Carden, Fred (2009). *Knowledge to Policy: Making the Most of Development Research.* Sage/IDRC. https://www.idrc.ca/en/book/knowledge-policy-making-most-development-research

Chambers, R. (2014). Perverse Payment by Results: Frogs in a Pot and Straitjackets for Obstacle Courses. Blog. https://participationpower.wordpress.com/2014/09/03/perverse-payment-by-results-frogs-in-a-pot-and-straitjackets-for-obstacle-courses/ (accessed 7 May 2015)

Court, J., Hovland, I. and Young, J. (2004). *Bridging Research and Policy in International Development: Evidence and the Change Process.* ITDG Bridging Research and Policy in International Development. Briefing Paper based on work conducted in the RAPID Programme at ODI. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/198.pdf

Count, J. and Young, J. (2003). *Bridging Research and Policy: Insights from 50 Case Studies.* ODI Working Paper 213. Overseas Development Institute. http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/180.pdf

Dagenais, C. *et al.* (2015). Collaborative Development and Implementation of a Knowledge Brokering Program to Promote Research Use in Burkina Faso, West Africa. *Global Health Action* 8.S1:1654-9880. http://www.globalhealthaction.net/index.php/gha/article/view/26004

Davies, P. (1999). What is Evidence-based Education? *British Journal of Educational Studies* 47(2):108-12. https://education.cuso.ch/fileadmin/education/document/Davies-1999-What-is-evidence-based-education.pdf

DFID (2008). *Working Paper Series: Research Communication*. UK DFID. https://assets.publishing.service.gov.uk/media/57a08b8eed915d622c000d57/ResearchStrategyWorkingPaperfinal_communications_P1.pdf

Douthwaite, B., S. Alvarez, K. Tehelen, D. Cordoba, G. Thiele and R. Mackay (2008). Participatory impact pathway analysis: A practical method for project planning and evaluation. In: *Fighting poverty through sustainable water use*. Proceedings of the CGIAR Challenge Program on Water and Food 2nd International Forum on Water and Food, Vol. 4, Addis Ababa, Ethiopia, 10-14 November 2008, eds. R.S. Bayot and E. Humphreys; 31. Colombo, Sri Lanka: CGIAR Challenge Program on Water and Food. https://cgspace.cgiar.org/bitstream/handle/10568/33649/11.5%20Participatory%20impact%20pathway%20analysis.pdf?sequence=1

Douthwaite, Boru, Sophie Alvarez, Graham Thiele, Ronald Mackay (2007). *Participatory impact pathways analysis*. In: Farmer First Revisited Conference (2007, Sussex, England). Papers presented. Future Agricultures Consortium; Institute of Development Studies (IDS), Sussex, GB. http://www.future-agricultures.org/farmerfirst/files/T1d_Douthwaite.pdf

Dutta, A. (2012). Deliberation, Dialogue and Debate: Why Researchers Need to Engage with Others to Address Complex Issues. *IDS Bulletin* 43.5, Brighton: IDS. http://bulletin.ids.ac.uk/idsbo/article/view/276

FHS (2014). *Increasing access to safe delivery*. Future Health Systems. http://futurehealthsys.squarespace.com/publications/increasing-access-to-safe-delivery.html

Fisher, C. and Vogel, I. (2008). *Locating the Power of In-Between: How Research Brokers and Intermediaries Support Evidence-Based Pro-poor Policy and Practice*. Brighton: IDS Knowledge Services, www.ids.ac.uk/files/dmfile/intconfpaper28Novwebsiteedit.pdf (accessed 7 May 2015)

Gaventa, J. (2006). Finding the Spaces for Change: A Power Analysis. *IDS Bulletin* 37.6: 23-33. http://www.powercube.net/wp-content/uploads/2009/12/finding_spaces_for_change.pdf

Georgalakis, James (2014). Influencing and engagement: Why let research programmes have all the fun? Blog: Research to Action. http://www.researchtoaction.org/2014/03/influencing-and-engagement-why-let-research-programmes-have-all-the-fun/

Greenhalgh, T. and Russell, J. (2006). Reframing Evidence Synthesis as Rhetorical Action in the Policy Making Drama. *Health Care Policy* 1(2):34-42, www.ncbi.nlm.nih.gov/pmc/articles/PMC2585323/ (accessed 7 May 2015)

Jones, H., Jones, N, Shaxson, L. and Walker, D. (2012). *Knowledge, Policy and Power in International Development*. Bristol: The Policy Press. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8201.pdf

Knezovich, J. (2014). ARCADE Presentation - What is a Policy Brief? Presentation delivered at the Institute of Development Studies. Open access resource. http://courses.arcade-project.org/mod/page/view.php?id=516

Kingdon, John W. (2003). *Agendas, alternatives, and public policies*. Boston: Little, Brown and Co.

Lindquist, E A (2001). *Discerning policy influence: framework for a strategic evaluation of IDRC - supported research*. University of Victoria, School of Public Administration, Victoria, BC, CA. https://idl-bnc.idrc.ca/dspace/handle/10625/29252

Lomas, J. (2006). Commentary: Whose Views Count in Evidence Synthesis? And When Do They Count? *Health Care Policy* 1.2: 55-57, www.ncbi.nlm.nih.gov/pmc/articles/PMC2585327/ (accessed 7 May 2015)

Lomas, J. (2005). Using Research to Inform Healthcare Managers' and Policy Makers' Questions: From Summative to Interpretive Synthesis, *Health Care Policy* 1.1: 55-71. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2585236/

Morton, Sarah (2015). Progressing research impact assessment: A 'contributions' approach. *Research Evaluation* Advance Access. August 14, 2015. http://rev.oxfordjournals.org/content/early/2015/08/14/reseval.rvv016.full.pdf+html

Morton, James, Louise Shaxson, Jonathan Greenland (2012). *Process Evaluation of the International Initiative for Impact Evaluation (2008-11): Final Report*. Overseas Development Institute and Triple Line Consulting. 27 September 2012. www.3ieimpact.org/media/filer_public/2013/01/07/3ie_proces_evaluation_2012_full_report_1.pdf

Newman, K.; Fisher, C. and Shaxson, L. (2012). *Stimulating Demand for Research Evidence: What Role for Capacity-building?* IDS Bulletin 43.5, Brighton: IDS. https://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/7524/IDSB_43_5_10.1111-j.1759-5436.2012.00358.x.pdf;jsessionid=C6AB582C08FFFC9EDA0A855AE0D49BD9?sequence=1

Pankhurst, J. (2011). Support Local Government to Get Research into Policy. SciDev.Net. www.scidev.net/global/communication/opinion/support-local-governance-to-get-research-into-policy-1.html (accessed 23 March 2015)

Perkins, Nick Ishmael Perkins and Simon Batchelor (2011). Learning about How to Communicate 'Learning' from and Within a Multi-sited Organisation. IDS Practice Paper in Brief. September 2011. IDS. http://www.ids.ac.uk/publication/learning-about-how-to-communicate-learning-from-and-within-a-multi-sited-organisation

Perkin, E. and Court, J. (2005). *Networks and Policy Processes in International Development: A Literature Review*. Working Paper 252. London: Overseas Development Institute. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/160.pdf

Peters, D.H.; Tran, N.T. and Adam T. (2013). *Implementation Research in Health: A Practical Guide*. Alliance for Health Policy and Systems Research, World Health Organization. www.who.int/alliance-hpsr/alliancehpsr_irpguide.pdf (accessed 28 January 2015).

Pope, C.; Mays, N. and Popay, J. (2005). Informing Policy Making and Management in Healthcare: The Place for Synthesis. *Health Care Policy* 1(2): 43-8. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2585321/

Porter, C. (2010). *What Shapes the Influence Evidence Has on Policy? The Role of Politics in Research Utilisation*. Young Lives Working Paper No. 62. Oxford: Young Lives. http://www.younglives.org.uk/content/what-shapes-influence-evidence-has-policy-role-politics-research-utilisation

Roe, E (1991). Development narratives or making the best of blueprint development. *World Development* 19(4). https://www.researchgate.net/publication/222215526_Development_Narratives_Or_Making_the_Best_of_Blueprint_Development

Roger, R (2006). A Decision-Maker's Perspectives on Lavis and Lomas. *Health Care Policy* 1(2):49-54. www.ncbi.nlm.nih.gov/pmc/articles/PMC2585335/ (accessed 23 March 2015)

Scott, A. and Munslow, T. (2015) *Tracking Research and Policy Conversations in Online Spaces*, IDS Evidence Report 122, Brighton: IDS. https://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/5921/ER122_TrackingResearchandPolicyConversationsinOnlineSpaces.pdf;jsessionid=380AB54C256CCA43405D842A46B07A8F?sequence=1

Scott-Villiers, P. (2012). This Research does not Influence Policy. *IDS Bulletin* 43.5, Brighton: IDS. https://opendocs.ids.ac.uk/opendocs/handle/123456789/7523

Shaxton, L (2012). Blog: Research Uptake and Impact: Are We in Danger of Overstating Ourselves? Impact Blog. London School of Economics. http://blogs.lse.ac.uk/impactofsocialsciences/2013/01/24/8864/ (accessed 23 March 2015)

STEPS Centre (undated). Participatory Impact Pathways Analysis (PIPA). http://steps-centre.org/methods/pathways-methods/vignettes/pipa/

Stone, Diane (2002). Using Knowledge: the dilemmas of 'Bridging Research and Policy'. *Compare: A Journal Of Comparative And International Education* 32(3). http://wrap.warwick.ac.uk/1799/1/WRAP_Stone_Compare.pdf

Stone, Diane with Simon Maxwell and Michael Keating (2001). Workshop: Bridging Research and Policy. Radcliffe House, Warwick University16-17 July 2001. http://www2.warwick.ac.uk/fac/soc/pais/research/researchcentres/csgr/research/keytopic/other/bridging.pdf

Sumner, A., Ishmael-Perkins, N. and Lindstrom, J. (2009). *Making Science of Influencing: Assessing the Impact of Development Research*. IDS Working Paper 335, Brighton: IDS. http://www.ids.ac.uk/publication/making-science-of-influencing-assessing-the-impact-of-development-research

Talbot, Colin and Carole Talbot (2014). Sir Humphrey and the professors: What does Whitehall want from academics? University of Manchester. http://www.policy.manchester.ac.uk/media/projects/policymanchester/1008_Policy@Manchester_Senior_Civil_Servants_Survey_v4(1).pdf

Vogel, Isobel (2012). *Review of the use of 'Theory of Change' in international development*. UK Department of International Development. https://assets.publishing.service.gov.uk/media/57a08a5ded915d3cfd00071a/DFID_ToC_Review_VogelV7.pdf

Wellcome Trust (2014). How can funders encourage more effective research uptake? Summary Report on an Exploratory Workshop. London 16 December 2014. http://www.ukcds.org.uk/sites/default/files/content/resources/Report%20on%20Exploratory%20Workshop%20on%20Research%20Uptake.pdf

Weiss, Carol H (1979). The many meanings of research Utilisation. *Public Administration Review* September/October 426-431. https://sites.ualberta.ca/~dcl3/KT/Public%20Administration%20Review_Weiss_The%20many%20meanings%20of%20research_1979.pdf

Weyrauch, V. and Diaz Langou, G. (2011). *Sound Expectations: From Impact Evaluations to Policy Change*. Working Paper 12. The International Initiative for Impact Evaluation (3ie): New Delhi. http://www.cippec.org/documents/10179/51827/PS+y+SC%2C%20SOUND++EXPECTATIONS+FROM+IMPACT+EVALUATIONS+TO+POLICY+CHANGE%2C%20Weyrauch+y+Diaz+Langou%2C%202011.pdf/2efea014-13f3-4135-9202-223f4f740982

Yaron and Shaxson (2008). *Good practice in evidence informed policy: an initial review for DFID*. http://webarchive.nationalarchives.gov.uk/20091005020131/http://www.dfid.gov.uk/Documents/publications/research/good-practice-policy-review.pdf

Young, John, Louise Shaxson, Harry Jones, Simon Hearn, Ajoy Datta and Caroline Cassidy (2014). *Rapid outcome mapping approach (ROMA): a guide to policy engagement and policy influence*. Overseas Development Institute: UK. http://www.roma.odi.org/diagnosing_the_problem.html#

End Notes

[i] However, the reader should take note that many research funding bodies will not be so open-minded and will only fund activities that fall within their own, potentially fairly narrow, definitions.

[ii] See www.health-policy-systems.com/series/systemsthinking for a collection of readings on complexity and www.coursera.org/learn/systems-thinking/ for an online course addressing this topic.

[iii] Note however that it will almost certainly not be possible to prevent both the individuals and the implementation team knowing which group they are in. Double or even single blinding will not be an option.

[iv] Note that confidence limits (and statistical significance tests) are most useful when sample sizes are relatively limited. With very large samples (~ 10,000) all estimates become extremely precise (and all test significant) (Lin et al. 2013).