



# Monitoring and Evaluating Digital Health Interventions

A practical guide to conducting  
research and assessment

Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment

ISBN 978-92-4-151176-6

© **World Health Organization 2016**

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (<http://www.wipo.int/amc/en/mediation/rules>).

**Suggested citation.** Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. Geneva: World Health Organization; 2016. Licence: CC BY-NC-SA 3.0 IGO.

**Cataloguing-in-Publication (CIP) data.** CIP data are available at <http://apps.who.int/iris>.

**Sales, rights and licensing.** To purchase WHO publications, see <http://apps.who.int/bookorders>. To submit requests for commercial use and queries on rights and licensing, see <http://www.who.int/about/licensing>.

**Third-party materials.** If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

**General disclaimers.** The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

Printed in Switzerland

Editing, design and layout by Green Ink, United Kingdom ([greenink.co.uk](http://greenink.co.uk))

Photos: Mark Leong for WHO/HRP, with exception of cover and page 117 (Carolyn Florey for United Nations Foundation)

# Contents

Acknowledgements .....	ii
Acronyms and abbreviations .....	iii
Preface .....	v
Executive summary .....	vi
Introduction .....	vii
<b>Chapter 1: Overview of monitoring and evaluation .....</b>	<b>1</b>
Part 1a: Defining goals for monitoring and evaluation .....	3
Part 1b: Developing an M&E plan for your digital health intervention .....	7
<b>Chapter 2: Setting the stage for monitoring and evaluation .....</b>	<b>13</b>
Part 2a: Articulating claims .....	14
Part 2b: Developing an M&E framework .....	22
Part 2c: Setting the stage: selecting indicators for digital health interventions .....	31
<b>Chapter 3: Monitoring digital health interventions .....</b>	<b>46</b>
Part 3a: Identifying stages of intervention maturity .....	50
Part 3b: Tools for monitoring .....	51
Part 3c: Digital health process monitoring components .....	54
<b>Chapter 4: Evaluating digital health interventions .....</b>	<b>65</b>
Part 4a: Key concepts for conducting digital health evaluations .....	67
Part 4b: Evaluation methods .....	78
Part 4c: Which evaluation activities are right for you? .....	89
<b>Chapter 5: Assessing data sources and quality for M&amp;E .....</b>	<b>96</b>
Part 5a: Introducing the data quality assessment approach and how to do it .....	98
Part 5b: Data quality assessment worksheet and instructions .....	100
Part 5c: Sample application of data quality assessment .....	108
<b>Chapter 6: Reporting your findings: the mHealth Evidence Reporting and Assessment (mERA) checklist .....</b>	<b>111</b>
Part 6a: How to use mERA .....	113
Part 6b: Methodological criteria .....	120
Annex I: Glossary .....	126

# Acknowledgements

This guide to monitoring and evaluating digital health interventions was developed through collaboration with the World Health Organization (WHO) Department of Reproductive Health and Research (RHR), the Johns Hopkins University Global mHealth Initiative (JHU-GmI) and the United Nations Foundation (UNF). This Guide was prepared by, in alphabetical order: Smisha Agarwal (JHU-GmI, Johns Hopkins School of Public Health [JHSPH]); Alain Labrique (JHU-GmI, JHSPH); Amnesty LeFevre (JHU-GmI, JHSPH); Garrett Mehl (WHO, RHR Department); Tigest Tamrat (WHO, RHR Department); Lavanya Vasudevan (Duke University); and Kelsey Zeller (JHSPH). Sean Broomhead and Tom Jones (African Centre for eHealth Excellence) provided critical inputs for Chapter 4 and co-wrote content on the economic evaluation of digital health interventions.

WHO would also like to thank Eliud Akama (Kenya Medical Research Institute [KEMRI]), Lakshmidurga Chava (Society for Elimination of Rural Poverty), Chika Hayashi (JHSPH), Kelly L'Engle (FHI 360), Larissa Jennings (JHSPH), Marc Mitchell (D-tree International), Thomas Odeny (KEMRI) and Trinity Zan (FHI 360) for their input and feedback on various sections of this Guide. Additional content on logical frameworks and other monitoring and evaluation components were drawn from programme materials provided by Anitha Moorthy and Karen Romano (Grameen Foundation); Sarah Andersson, Yasmin Chandani, Megan Noel and Mildred Shieshia (John Snow, Inc.); Marcha Bekker and Debbie Rogers (Praekelt Foundation); and Jesse Coleman (Wits Reproductive Health and HIV Institute).

The meaningful contributions of a diverse group of implementers from the digital health space, many of whom have been supported by the Innovation Working Group (IWG) Catalytic mHealth Grant Mechanism, have been essential to the development of this Guide. This grant mechanism, which consists of a collaboration between the UNF and the WHO Department of Reproductive Health and Research, including the UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), has assisted 26 mHealth projects in the process of scaling up by providing funding, technical assistance and joint learning opportunities over the past four years.

The authors are additionally grateful to the individuals who provided useful suggestions during the preparation of earlier drafts, including Carolyn Florey (UNF), Francis Gonzales (UNF), Michelle Hindin (WHO, RHR Department) and Abigail Manz (UNF). We also thank Jane Patten (editor), Jeff Walker and Eszter Saródy (designers) who provided feedback and creative input during the review stage, on behalf of Green Ink, United Kingdom.

Finally, the authors are grateful to the Norwegian Agency for Development Cooperation (Norad), and wish to extend particular thanks to Helga Fogstad and Haitham el-Noush for their support to the IWG Catalytic mHealth Grant Mechanism, and their leadership and vision for the use of mobile health technologies to improve reproductive, maternal, newborn and child health.

# Acronyms and abbreviations

<b>ANC</b>	Antenatal care
<b>CBA</b>	Cost–benefit analysis
<b>CCA</b>	Cost–consequence analysis
<b>CEA</b>	Cost–effectiveness analysis
<b>CHW</b>	Community health worker
<b>CIEL</b>	Columbia International eHealth Laboratory
<b>CONSORT</b>	Consolidated Standards of Reporting Trials
<b>CMA</b>	Cost-minimization analysis
<b>CUA</b>	Cost–utility analysis
<b>DALY</b>	Disability-adjusted life year
<b>DHIS</b>	District Health Information System
<b>DHS</b>	Demographic and Health Survey
<b>eHealth</b>	Electronic health
<b>FGD</b>	Focus group discussion
<b>HEI</b>	HIV-exposed infant
<b>HIPAA</b>	Health Insurance Portability and Accountability Act
<b>HIS</b>	Health information system
<b>HIV</b>	Human immunodeficiency virus
<b>HL7</b>	Health Level 7 (data standard)
<b>HMIS</b>	Health management information system
<b>HRP</b>	The UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction
<b>ICD</b>	International Classification of Diseases
<b>ICT</b>	Information and communications technology
<b>IDI</b>	In-depth interview
<b>IR</b>	Immediate results
<b>ISO</b>	International Organization for Standardization
<b>IT</b>	Information technology
<b>IVR</b>	Interactive voice response
<b>IWG</b>	Innovation Working Group
<b>JHU-GmI</b>	Johns Hopkins University Global mHealth Initiative
<b>JHSPH</b>	Johns Hopkins University School of Public Health
<b>JSI</b>	John Snow, Inc.
<b>K4Health</b>	Knowledge4Health
<b>KEMRI</b>	Kenya Medical Research Institute

<b>M&amp;E</b>	Monitoring and evaluation
<b>MAMA</b>	Mobile Alliance for Maternal Action
<b>MAPS</b>	mHealth Assessment and Planning for Scale
<b>mERA</b>	mHealth Evidence Reporting and Assessment
<b>mHealth</b>	The use of mobile and wireless technologies for health
<b>MICS</b>	Multiple Indicator Cluster Survey
<b>MNO</b>	Mobile network operator
<b>MOH</b>	Ministry of health
<b>MOTECH</b>	Mobile Technology for Community Health (Ghana)
<b>N/A</b>	Not applicable
<b>NGO</b>	Nongovernmental organization
<b>Norad</b>	Norwegian Agency for Development Cooperation
<b>OpenMRS</b>	Open Medical Record System
<b>PAR</b>	Participatory action research
<b>PMTCT</b>	Prevent mother-to-child transmission (of HIV)
<b>PNC</b>	Postnatal care
<b>PRISM</b>	Performance of Routine Information System Management
<b>RCT</b>	Randomized controlled trial
<b>RMNCAH</b>	Reproductive, maternal, newborn, child and adolescent health
<b>RMNCH</b>	Reproductive, maternal, newborn and child health
<b>SBA</b>	Skilled birth attendant
<b>SMART</b>	Specific, measurable, attainable, relevant and time-bound
<b>SMS</b>	Short messaging service (also known as text messages)
<b>SOP</b>	Standard operating procedure
<b>SP</b>	Sulfadoxine-pyrimethamine, used in preventive treatment of malaria in pregnancy
<b>SRS</b>	Software requirements specification
<b>STI</b>	Sexually transmitted infection
<b>STROBE</b>	STrengthening the Reporting of OBservational studies in Epidemiology
<b>UHC</b>	Universal health coverage
<b>UI</b>	User interface
<b>UNDP</b>	United Nations Development Programme
<b>UNF</b>	United Nations Foundation
<b>UNFPA</b>	United Nations Population Fund
<b>UNICEF</b>	United Nations Children's Fund
<b>USAID</b>	United States Agency for International Development
<b>USSD</b>	Unstructured supplementary service data
<b>WHO</b>	World Health Organization

# Preface

Over the past five years, substantial progress has been made in defining terms around the landscape of digital, mobile and wireless technologies for health, or digital health – also commonly referred to as mHealth or eHealth. Broadly, digital tools are increasingly being tested, evaluated and, in some instances, integrated at scale into health systems in low- and middle-income countries striving to meet goals of universal health coverage (UHC). Along with the proliferation of small innovation projects testing the use of mobile and digital technologies, concerted efforts to harmonize and learn from these deployments are also under way.

Since 2011, in partnership with the World Health Organization (WHO) Department of Reproductive Health and Research (RHR), the United Nations Foundation (UNF) has been supported by the Norwegian Agency for Development Cooperation (Norad) to oversee three yearly rounds of grants to mHealth projects. A total of 26 organizations received financial investments and technical assistance towards the goal of demonstrating potential for scaling up digital health innovations to catalyse achievement of the health-focused United Nations Millennium Development Goals (MDGs). The research and technical support provided through this mechanism, with assistance from the Johns Hopkins University Global mHealth Initiative (JHU-Gml), have afforded numerous opportunities to engage with and learn from implementing partners on the ground, across Asia and Africa.

This resource represents the collective learning from five years of engagement with agencies working to strengthen their digital health deployments, develop robust evaluations, and scale up their activities nationally and regionally. The lessons learnt from working with these partners are described in this document, which provides high-level guidance and systematic direction to programme planners and implementers embarking on similar journeys. Specifically, this Guide provides an introduction to the approaches and methods that were identified as useful for (i) the monitoring of project (i.e. intervention) deployments, focusing on the quality and fidelity of the intervention inputs; and (ii) the evaluation of project outputs and impacts across a number of axes, from user satisfaction to process improvements, health outcomes and cost-effectiveness.

Although more in-depth texts and curricula are available on the methods discussed, this Guide focuses on presenting pragmatic highlights and experience-informed tips for implementers to consider, together with links and resources for further study. It leads the reader through the development of value “claims”, evaluation designs and indicators associated with their digital health intervention, an assessment of the quality and availability of the data from their intervention, and finally, a series of guidelines for the reporting of findings.



# Executive summary

This Guide provides step-wise guidance to improve the quality and value of monitoring and evaluation (M&E) efforts in the context of digital health interventions, which are also commonly referred to as mHealth or eHealth interventions. Among the many challenges identified in the digital health landscape, those of programme monitoring and impact evaluation remain areas of ongoing exploration. Digital health interventions are often very dynamic, evolving through several stages of maturity during which the M&E needs of the intervention are also changing rapidly. Digital health intervention projects typically begin with exploring basic questions of whether the intervention addresses the identified needs, including technical functionality and feasibility, followed by assessment of user satisfaction, then move towards efforts to evaluate the effectiveness, attributable impact and, ultimately, “value for money” of the intervention.

The Guide assists the reader to navigate through the development of value “claims”, the selection of indicators and evaluation designs associated with their digital health interventions, as well as approaches for the assessment of the quality and availability of the data from their interventions, and finally, guidelines for the reporting of findings. This progression of activities requires a combination of methods, both qualitative and quantitative, to answer the questions being asked about digital health interventions. Accordingly, this resource directs the reader through a journey that begins with defining the basic technical requirements and continues to early implementation testing and monitoring, through to the evaluation and reporting of intervention impact.



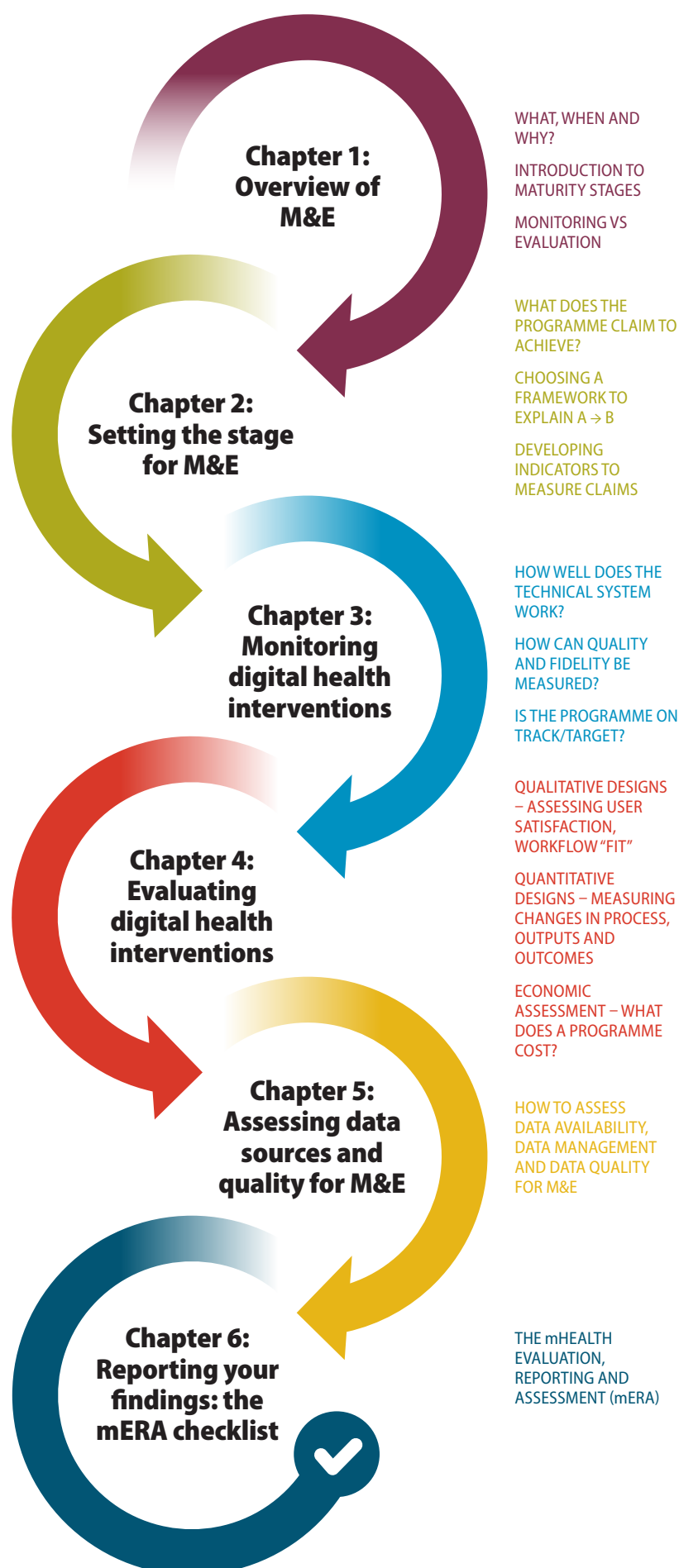
# Introduction

This Guide is structured to guide the reader through the pathway described in Figure 1, beginning with a broad overview in Chapter 1 to describe the goals for monitoring and evaluation (M&E), explicitly distinguishing between the efforts aimed at monitoring implementations and those aimed at evaluating their impact. Chapter 2 guides the reader to formulate specific intervention claims and develop indicators specific to those claims, including the selection of process indicators that reflect implementation fidelity. Additionally, Chapter 2 introduces readers to the selection and development of a framework to guide the intervention assessment. Once a framework, claims and indicators have been developed and established, Chapter 3 takes readers through the set-up of a monitoring plan, focusing on technical stability and performance. In Chapter 4, we shift to the realm of evaluation, to introduce the reader to qualitative, quantitative and economic methods commonly used to generate data in support of programme claims.<sup>1</sup>

Some readers may be using the Guide late in their implementation process, in which case the scope for generating new data or introducing new evaluation methods may be limited – these readers can skip ahead to Chapter 5, which focuses on methods for assessing, and improving, the quality of data being collected. Reviewing the data sources is critical, since poor-quality data can undermine both monitoring and evaluation efforts.

The last part of the Guide, Chapter 6, focuses on reporting findings from the programme, an often neglected, but critical area – decision-makers look to these findings for support when seeking to invest in digital health strategies. To date, inconsistent or incomplete reporting

**Figure 1. The Guide Roadmap: six steps for the M&E of digital health interventions**



<sup>1</sup> Please see the glossary for definitions of terms; chapters also include definitions boxes for terms that are central to the topic of each chapter.

of digital health interventions remains a major barrier to the synthesis of evidence in support of particular strategies. For governments, donors and multilateral agencies to appreciate the potential impact of a digital health intervention, complete and robust reporting of individual intervention projects is vital.

The Guide makes a distinction between steps intended to **monitor** implementation activities – that is, to assure fidelity, quality and coverage of the intervention being delivered to a population – and those intended to **evaluate** programme activities – that is, to attribute some output, outcome or economic value to the intervention.

Although these efforts are often closely intertwined during implementation, conceptually it is simpler to disentangle them in the planning stage. This allows programme managers to focus separately on establishing systems that measure and monitor how consistently a programme is implementing its planned activities and meeting its objectives, understanding that this feeds into a broader evaluation agenda of understanding the impact of the programme and whether or not it has achieved its goal.

## Intended audience

This Guide is intended for implementers and researchers of digital health activities, as well as policy-makers seeking to understand the various stages and opportunities for systematically monitoring implementation fidelity and for evaluating the impact of digital health interventions.

At the start of this Guide, we make the assumption that you, the reader, have already embarked on your digital health journey and completed the requisite groundwork for implementing your digital health intervention, from problem analysis to user-centred design, guided by tools such as K4Health's *mHealth planning guide* (1) and the *MAPS Toolkit* (2).



**Digital health:** The use of digital, mobile and wireless technologies to support the achievement of health objectives. Digital health describes the general use of information and communications technologies (ICT) for health and is inclusive of both mHealth and eHealth.

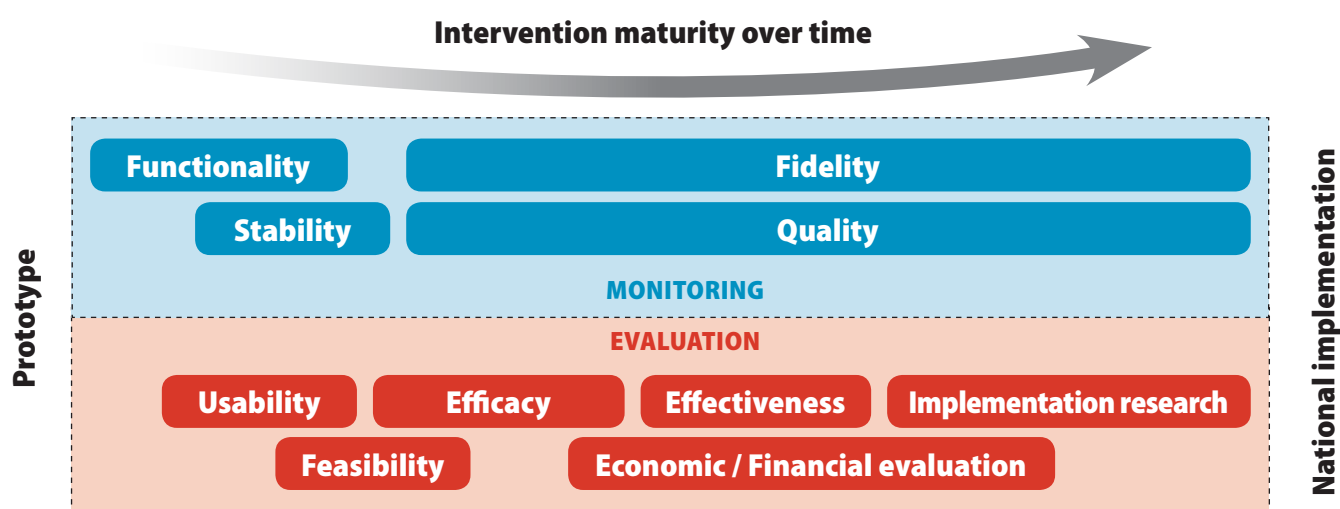


# Chapter 1: Overview of monitoring and evaluation



There is broad consensus that a common framework for evaluating digital health interventions<sup>2</sup> is vital to generate evidence required for decision-making on the appropriate approach to integrate effective strategies into broader national health systems. Careful monitoring and systematic evaluations of digital health interventions, however, have been few in number, in contrast to the proliferation of digital health pilot projects addressing various health needs in low- and middle-income countries. In recent years, as governments and donors have increased the level of scrutiny imposed on these innovations, calls for better assessment of the quality and impact of these intervention projects have arisen. Within the recently published WHO *MAPS Toolkit: mHealth Assessment and Planning for Scale*, robust monitoring and evaluation plans were specifically identified as essential to support potential intervention scale-up (2).

**Figure 1.1. Intervention maturity life-cycle schematic, illustrating concurrent monitoring (blue/upper) and evaluation (red/lower) activities that occur as an intervention matures over time (left to right) from a prototype application to national implementation**



However, as new digital health interventions emerge, they commonly undergo what is recognized as an intervention maturity life-cycle, depicted in Figure 1.1, as they journey from prototype of the digital health system towards possible national-level implementation of the digital health intervention. During this life-cycle, concurrent monitoring and evaluation activities should be planned, often in parallel, supporting each other. As the intervention matures, the M&E needs will evolve – from monitoring the system’s technical functionality and stability, towards continuous, real-time monitoring of its consistency in producing the expected outputs, at a pre-defined level of quality. The evaluation of the digital health system and intervention over time is an attempt to attribute a range of outcomes to the technology-based intervention – from assessing how easily end-users can interact with the system (usability), to the health impacts attributed to the intervention (efficacy/effectiveness), to the affordability of the system (economic/financial evaluation). In later stages of maturity, questions may arise around the integration of the system and its data streams within the broader health system architecture and policy environment, as interventions attempt to reach and sustain national scale (implementation science).

This chapter provides a general overview of fundamental considerations to be reviewed when conceptualizing and embarking on M&E activities for digital health interventions. By clarifying the differences and linkages between monitoring and evaluation, this chapter addresses key issues of articulating the overall goals and intentions for the M&E efforts. This chapter also underlines the appropriateness of different M&E questions to be asked throughout the life-cycle (stages of maturity) of a digital health intervention. This first chapter concludes by guiding readers in their development of a concrete plan to execute the envisioned M&E activities, which are detailed in subsequent chapters.

<sup>2</sup> “Intervention” in this Guide can also refer to projects, programmes, initiatives and other activities that are being monitored and evaluated.

# Part 1a: Defining goals for monitoring and evaluation



## HOW WILL THIS SECTION HELP ME?

This section will:

- ✓ Help you to determine the overall goal of M&E and the needs for M&E activities, so that you can strategically direct resources and efforts.
- ✓ Guide you, by distinguishing between monitoring and evaluation, to identify the relevant needs and pose appropriate questions.
- ✓ Highlight illustrative research questions to be asked over the course of the digital intervention's lifecycle and stage of maturity.

## What is monitoring?

**Process monitoring** is generally defined as the continuous process of collecting and analysing data to compare how well an intervention is being implemented against expected results (3). In this Guide (i.e. in the context of digital health interventions), “monitoring” and “process monitoring” are used interchangeably to refer to the routine collection, review and analysis of data, either generated by digital systems or purposively collected, which measure implementation fidelity and progress towards achieving intervention objectives.

The six stages of the intervention maturity life-cycle, as represented in Box 1.1, help to illustrate how the levels of inquiry “graduate” from a focus on the technical (or device/system) factors to the interaction between the user and that system, eventually introducing more complex questions around the system’s performance within a health system context and at various levels of scale. Stage 1 and 2 M&E questions focus on the technology itself, as illustrated on the left-hand side of Box 1.1. Stage 3 questions relate to the interface between the end-user and the technology. In Stage 4, limited deployments aim to measure attributable impact on specific processes or outcomes, usually in controlled environments. Stage 5 and 6 deployments are gradually at larger levels of scale, testing effectiveness in non-research settings, without tight controls on the delivery of the intervention, aiming to measure cost and cost-effectiveness, or identify challenges to scale-up in the realm of policy changes or organizational change management.



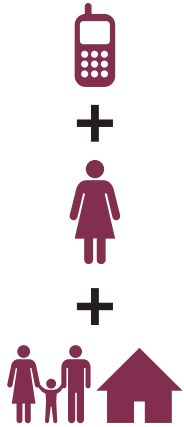
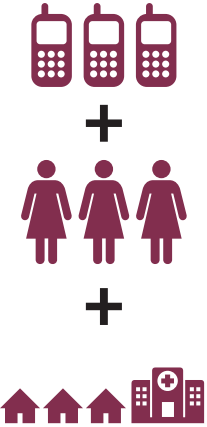

Overall, monitoring activities should be answering this question: Is the intervention working as it was intended? Monitoring activities can measure changes in performance over time, increasingly in real time, allowing for course-corrections to be made to improve implementation fidelity.

Plans for monitoring of digital health interventions should focus on generating data to answer the following questions, where “system” is defined broadly as the combination of technology software, hardware and user workflows:

- Does the system meet the defined technical specifications?
- Is the system stable and error-free?
- Does the system perform its intended tasks consistently and dependably?
- Are there variations in implementation across and/or within sites?
- Are **benchmarks** for deployment being met as expected?

Effective monitoring entails collection of data at multiple time points throughout a digital health intervention’s life-cycle and ideally is used to inform decisions on how to optimize content and implementation of the system. As an iterative process, monitoring is intended to lead to adjustments in intervention activities in order to maintain or improve the quality and consistency of the deployment.

**Box 1.1. Schematic depiction of the six stages of the intervention maturity life-cycle from pre-prototype to national-level deployment**

Stage of maturity	1 & 2: Pre-prototype/prototype	3: Pilot	4: Demonstration	5: Scale-up	6: Integration/sustainability
					
<b>Monitoring goals</b>	Functionality, stability	Fidelity, quality			
<b>Stages of evaluation</b>	Feasibility/usability	Efficacy	Effectiveness	Implementation science	
<b>Illustrative number of system users</b>	10–100	100–1000	10 000+	100 000+	
<b>Illustrative measurement targets</b>	<ul style="list-style-type: none"> <li>■ Stability (system uptime/failure rates)</li> <li>■ Performance consistency</li> <li>■ Standards adherence (terminology, interoperability, security)</li> </ul>	<ul style="list-style-type: none"> <li>■ User satisfaction</li> <li>■ Workflow “fit”</li> <li>■ Learning curve (design)</li> <li>■ Cognitive performance/errors</li> <li>■ Reliability</li> </ul>	<ul style="list-style-type: none"> <li>■ Changes in process (time to X)</li> <li>■ Changes in outcome (system performance/health)</li> </ul>	<ul style="list-style-type: none"> <li>■ Changes in process/outcome in less controlled environment</li> <li>■ Reduction of cost</li> <li>■ Total cost of implementation</li> <li>■ Error rates</li> <li>■ Learning curve of users</li> </ul>	<ul style="list-style-type: none"> <li>■ Improvements in coverage</li> <li>■ Changes in policy, practices attributable to system</li> <li>■ Extendability to new use-cases</li> <li>■ Adaptability to other cadres of users</li> <li>■ Health impact</li> </ul>

## What is evaluation?

**Evaluation** is generally defined as the systematic and objective assessment of an ongoing or completed intervention with the aim of determining the fulfilment of objectives, efficiency, effectiveness, impact and sustainability (3). Evaluation, in this Guide, refers to measures taken and analysis performed in order to assess (i) the interaction of users or a health system with the digital health intervention strategy, or (ii) changes attributable to the digital health intervention. Whereas monitoring (defined above) focuses on measuring properties that are intrinsic (inward) to the digital health system or intervention, evaluation concentrates instead on metrics that are extrinsic (outward) to the intervention. Ideally, the intention is to demonstrate attribution – that is, to show that the changes in these extrinsic metrics have occurred as a result of the digital health intervention.

Monitoring begins with the measurement of usability, focusing on the quality of the interaction between the user and the technology, and feasibility, which explores contextual readiness, ranging from human resource capacity to the technical ecosystem (e.g. connectivity, electrical grid stability, mobile phone access). Once established, the challenge of measuring the extent to which any observed changes in outcome and impact can be attributed to the digital health intervention begins.

Attributing change to the intervention is one of the most difficult challenges, and is addressed by a combination of the research method selected, the quality of the data collected and the appropriateness of the comparison, or



#### KEY TERMS

**Benchmark:** Reference point or standard against which performance or achievements can be assessed (3).

**Evaluation:** The systematic and objective assessment of an ongoing or completed intervention with the aim to determine the fulfilment of objectives, efficiency, effectiveness, impact and sustainability (3). In this Guide (i.e. in the context of digital health interventions), evaluation is used to refer to measures taken and analysis performed to assess (i) the interaction of users or a health system with the digital health intervention strategy, or (ii) changes attributable to the digital health intervention.

**Process monitoring:** The continuous process of collecting and analysing data to compare how well an intervention is being implemented against expected results (3). In this Guide (i.e. in the context of digital health interventions), “**monitoring**” or “**process monitoring**” are used interchangeably to refer to the routine collection, review and analysis of data, either generated by digital systems or purposively collected, which measure implementation fidelity and progress towards achieving intervention objectives.

counterfactual. Evaluation plans for digital health interventions should focus on generating data that can be used as a basis for assessing whether observed changes in behaviour, processes or health outcomes can be attributed to the intervention. A combination of the following questions (which are illustrative but not comprehensive) can be used for measuring attribution:

- Usability
  - ✓ Is the digital health system usable by the targeted end-user(s), and does it fit within their workflow?
  - ✓ How steep is the learning curve before a user can demonstrate proficient system use?
  - ✓ What are the rates of error – in using the system or in workflows – as a result of system use/misuse?
- Efficacy
  - ✓ Has the digital health intervention changed processes (e.g. time between event X and response Y) in a research setting?
  - ✓ Has the digital health intervention changed outcomes (e.g. worker performance, such as guideline adherence, or patient health outcomes) in a research setting?
- Effectiveness
  - ✓ Has the digital health intervention changed processes (e.g. time between event X and response Y) in a non-research setting?
  - ✓ Has the digital health intervention changed outcomes (e.g. worker performance, such as guideline adherence, or patient health outcomes) in a non-research setting?
- Cost
  - ✓ Has the digital health intervention reduced costs associated with the delivery of health services?
  - ✓ Has the digital health intervention introduced costs that are commensurate with benefits provided?

## Linking monitoring and evaluation

**“Evaluation asks whether the project is doing the right things, while monitoring asks whether the project is doing things right.” – Pritchett et al., 2013 (4)**

Monitoring and evaluation activities occur in close complement to each other. For clarity’s sake, we introduce them as distinct, albeit intertwined, streams of activities in this Guide. Evaluation strategies build on monitoring data and implementation activities to measure and attribute changes in the health system (or impact on clients) occurring as a result of the intervention. The schematic in Box 1.1 illustrates this interrelationship between these two domains of inquiry.



Poorly implemented interventions lacking robust monitoring activities are unlikely to generate the impact expected from them. There is often a tendency to assume that a digital health intervention was not effective, even though evaluation results may be based on poor monitoring of the implementation. For example, having a high proportion of clients who miss the text messages due to connectivity issues could yield evaluation results indicating that text message reminders did not improve uptake of the intervention. The M&E team may conclude that the text messages were ineffective, but this would be the wrong conclusion since the reminders cannot improve uptake if those reminders have not been received. However, this lack of rigorous monitoring leads to an inability to appropriately state whether an intervention's ineffectiveness is directly due to the intervention (e.g. it doesn't work) or a result of the implementation.

# Part 1b: Developing an M&E plan for your digital health intervention



## HOW WILL THIS SECTION HELP ME?

This section will:

- ✓ Identify concrete steps for designing and planning the evaluation of a digital health intervention.
- ✓ Help you determine which components of the intervention should be evaluated.
- ✓ Guide you in developing a plan of action for executing defined M&E objectives, and understanding the resources and effort required to carry out the activities.

This Guide proposes a seven-step approach to designing M&E activities for digital health interventions. Each step is introduced in Figure 1.2 and outlined in the text that follows.

**Figure 1.2. A pathway for monitoring and evaluating digital health interventions**

<b>STEP 1</b>	Define the stage of maturity, stage of evaluation, and appropriate claims
<b>STEP 2</b>	Develop an underlying framework
<b>STEP 3</b>	Identify evidence needs and evaluation objectives
<b>STEP 4</b>	Finalize a study design
<b>STEP 5</b>	Determine who will carry out monitoring and evaluation activities
<b>STEP 6</b>	Timing and resources
<b>STEP 7</b>	Define an M&E implementation plan

## Step 1. Define the stage of maturity, stage of evaluation, and appropriate claims

A critical first step to defining an appropriate approach to evaluating a digital health intervention lies in appropriately classifying (a) where the technology is in terms of stage of maturity, (b) which stage of evaluation corresponds to the intervention and (c) which **claims** are appropriate (see Box 1.2).



**Claim:** A statement of anticipated benefits of the digital health system or intervention.

- Stage of maturity:** The stages of maturity span across the continuum from pre-prototype, through prototype, pilot, and demonstration, to scale-up and, ultimately, integrated and sustained implementations (see Box 1.1). Project teams must first agree on where the digital health intervention is situated along this continuum in order to determine the appropriate evaluation activities and avoid embarking on premature assessments.
- Stage of evaluation:** The stage of evaluation invariably corresponds to the stage of maturity. The stages of evaluation include assessments to determine feasibility, usability, efficacy, effectiveness, or assessment of the implementation factors to improve the likelihood of achieving a successful integrated and sustained implementation. These stages are elaborated further in Chapter 4, which focuses on evaluation.

- c. **Appropriate claims:** To better understand the claims that it is necessary or possible to make for a given digital health intervention, and to guide you in defining your evaluation approach, two main questions must be considered:
- Are we evaluating the added benefit of the digital health component to optimize the delivery of an existing or already validated health intervention? (e.g. Do digital supply chain systems improve the coverage of childhood vaccines?) and/or
  - Are we evaluating the effectiveness of the digital health intervention to directly and independently trigger a health outcome (i.e. where the effectiveness is not yet known)? (e.g. Do electronic decision support systems improve the quality of services provided by health-care providers?)

See Chapter 2, Part 2a, for more information on developing appropriate claims.

### Box 1.2. What kind of evaluation do we need, based on where we are now?

**Stage of maturity:** Is the digital health intervention being developed and evaluated for the first time, or is it mature and undergoing scale-up?

**Stage of evaluation:** Is the digital health intervention being evaluated to determine whether the system functions, is effective, or is able to undergo scale-up?

**Claims:** Is the digital health intervention being evaluated with the intention of improving delivery of an intervention with known efficacy, such as a vaccine or drug? Or is the digital health intervention itself novel and its ability to improve health outcomes still unknown?

Table 1.1 links the taxonomic stages of maturity (a) with the stages of evaluation (b), as well as corresponding claims or the broader aims for each stage (c). In Chapter 2, Part 2a, claims are covered in more detail along with linkages to broader study objectives and aims.

## Step 2. Develop an underlying framework

To guide and support the M&E activities, you need first to develop an underlying framework. Frameworks outline the process and rationale to guide you towards achievement of your research goals. Defining a framework will help you to (i) define and understand the objectives of the intervention; (ii) conceptualize the relationship between these different objectives; (iii) define the underpinning project activities required to achieve your goals and objectives; and (iv) describe the anticipated outcomes.

In Chapter 2, Part 2b, the Guide defines and outlines some of the most commonly used types of frameworks: (i) Conceptual frameworks; (ii) Results frameworks; (iii) Theory of change frameworks; and (iv) Logical frameworks. Deciding which type of framework is most relevant for you will depend on key stakeholder needs and project context and complexity. Ultimately, adoption of a framework will strengthen the design, implementation, and M&E of your digital health intervention. Ideally they are developed through a consultative process, and revised throughout the life of a project in response to early M&E data, changes in assumptions and/or project design/implementation.

## Step 3. Identify evidence needs and evaluation objectives

Where goals provide a broad statement about the desired long-term outcomes and impact of your project, objectives are a statement of Specific, Measurable, Attainable, Relevant and Time-bound (SMART) results. Objectives should be defined through a collaborative process with key stakeholders by first reviewing the broader project goals and anticipated outcomes. Outcomes should be measurable using indicators, and should be defined to facilitate the generation of evidence required as a basis for key decision-making. Finally, objectives should be linked with the timing and stage of evaluation (see Box 1.3). SMART objectives are further described in Chapter 2, Part 2c.

**Table 1.1. Linking stages of maturity with evaluation methods and claims**

	Stage of maturity	Stage of evaluation	Claim
	(a)	(b)	(c)
<b>Early</b>	<p><b>Pre-prototype:</b> This stage includes hypothesis building, needs/context assessment, and testing of usability/feasibility and technical stability.</p> <p><b>Prototype:</b> During this phase, user-focused designs are created and tested, and functionality, technical stability and usability are tested in an iterative process. Ways to improve the system are examined to enhance relevance.</p> <p><b>Pilot:</b> This stage examines whether the digital health intervention can produce the desired effect under controlled circumstances. The pilot project is usually a single deployment.</p>	<p><b>Feasibility:</b> Assess whether the digital health system works as intended in a given context.</p> <p><b>Usability:</b> Assess whether the digital health system can be used as intended by users.</p> <p><b>Efficacy:</b> Assess whether the digital health intervention can achieve the intended results in a research (controlled) setting.</p>	<p><b>Technology:</b> Prototypes are functional and usable.</p> <p>Feasibility testing demonstrates end-user acceptance and expected data integrity and validity.</p> <p><b>Intervention:</b> Implementation protocols are utilized as intended by users.</p> <p><b>Technology:</b> Technology withstands testing under optimal field circumstances.</p> <p><b>Health:</b> Health improvements (outputs/outcomes/impact) demonstrated on a small scale, under optimal circumstances, warranting further testing.</p>
<b>Mid</b>	<p><b>Demonstration:</b> In this stage, the intervention is no longer taking place in controlled conditions but is still limited in terms of population/geography (usually restricted to a particular region or subregion).</p> <p>This stage seeks to understand the costs and implementation requirements needed to both deliver the intervention at high fidelity and replicate the uptake in new contexts.</p>	<p><b>Effectiveness:</b> Assess whether the digital health intervention can achieve the intended results in a non-research (uncontrolled) setting.</p>	<p><b>Health services delivery</b> at moderate-scale implementation in a non-research setting is determined to be:</p> <ul style="list-style-type: none"> <li>■ feasible</li> <li>■ high quality</li> <li>■ cost-effective</li> <li>■ improving the effectiveness of bringing about positive change in health outcomes.</li> </ul>
<b>Advanced</b>	<p><b>Scale-up:</b> In this stage, approaches are ready to be optimized and scaled up across multiple subnational, national or population levels.</p> <p><b>Integrated and sustained programme:</b> Efforts at this stage are focused on determining the necessary components of an enabling environment that will support impact of the intervention at a large scale (i.e. policies, financing, human resources, interoperability, etc.). The intervention has been integrated into a broader health system.</p>	<p><b>Implementation science:</b> Assess the uptake, integration and sustainability of evidence-based digital health interventions for a given context, including policies and practices.</p>	<p><b>Technology:</b> Technology is functional and being effectively implemented at scale.</p> <p>Support systems are in operation to ensure continuous service provision.</p> <p><b>Health services delivery</b> at large-scale implementation through integrated service delivery is determined to be:</p> <ul style="list-style-type: none"> <li>■ feasible</li> <li>■ high quality</li> <li>■ cost-effective</li> <li>■ improving the effectiveness of bringing about positive change in health outcomes.</li> </ul>

### Box 1.3. Defining M&E objectives

1. Identify the key stakeholders
2. Discuss with implementers, funders and other key stakeholders
3. Review project goals and anticipated outcomes
4. Identify the evidence required to influence future decision-making
5. Draft objectives that correspond with the appropriate stage of maturity and evaluation
6. Ensure objectives are SMART: specific, measurable, attainable, relevant and time-bound.

## Step 4. Finalize a study design

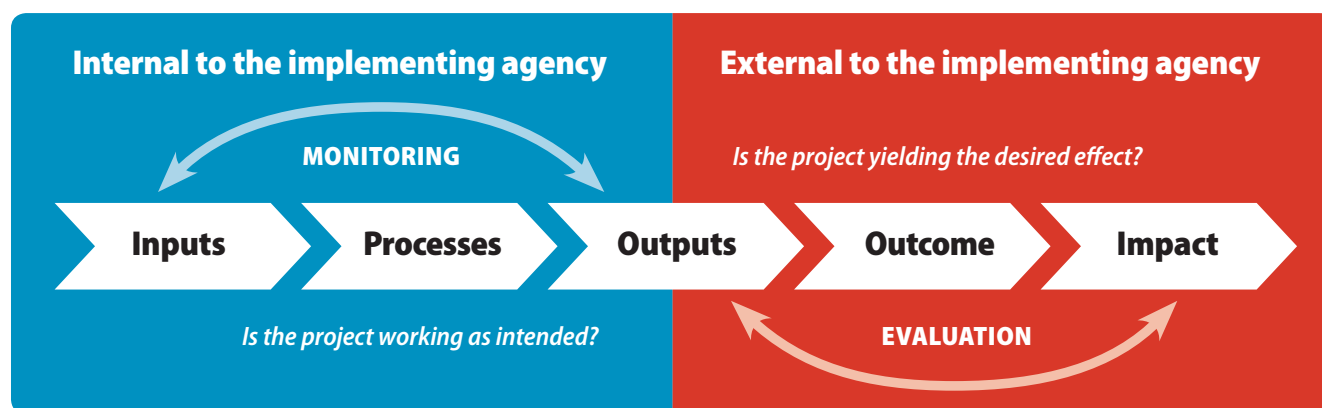
Once you have developed a framework and articulated the evidence needs, you need to decide on the optimal study design appropriate for the implementation, monitoring and evaluation of your project. The study design selected will help inform decision-making on evidence generation and the scope of M&E activities. Study design considerations should be determined by the stage of evaluation within which a given digital health intervention falls, and should take into account evidence hierarchies. Chapter 4 expands on these terms and describes various evaluation methods.

## Step 5. Determine who will carry out monitoring and evaluation activities

When planning your evaluation, you need to consider who will carry out the M&E activities. Internal evaluations may sometimes be perceived as lacking independence. Often, the evaluators are affiliated with the implementers, and this may create a conflict of interest and influence the evaluation results if the results are tied to funding for the project. However, internal evaluations may be less expensive, and if done in a rigorous manner they can still answer critical research questions. External evaluations are carried out by an individual or institution that is independent from the project and its implementers and, as a result, are considered to retain a degree of impartiality, which imparts a higher level of credibility on the evaluation results. However, these evaluations are more costly and may require additional time to get the research partner on board.

For many digital health interventions, monitoring will be carried out internally by the implementing agency and focus on linkages between inputs, processes and outputs. In contrast, evaluation efforts to determine an intervention's effect on health outcomes and impact may be conducted by a research organization external to the project and its intended clients or beneficiaries (see Figure 1.3).

**Figure 1.3. Schematic diagram of the interplay between monitoring and evaluation activities**



Source: adapted from Pritchett et al. 2013 (4).

## Step 6. Timing and resources

The process of designing evaluations is an iterative process, in which consideration of timing and available resources inform the refinement of objectives formulated in Step 3 to ensure their feasibility.

With regard to timing, the design of evaluations must take into consideration where in the life-cycle of project development and implementation a given digital health intervention is (or will be) at the inception of evaluation activities. For example, the range of available evaluation options will be more limited if a given digital health intervention is already midway into implementation compared to those available if plans for evaluation and the evaluation activities were initiated prior to project implementation. Evaluation may take place at the following points in time: (i) at the inception of a project (prospective); (ii) following the project's initiation/introduction; or (iii) at the project's completion (retrospective). Prospective evaluations are preferred.

In addition to considerations related to the start of the evaluation, the time available to carry it out and the time needed to demonstrate results must be weighed. Finally, the available resources – financial, human and physical (supplies/equipment) – must also be quantified. While it is recommended that 10% of the total budget available for project implementation be allocated to support evaluation activities (5), this might not be feasible or adequate in practice. In some instances, more resources may be needed for an evaluation that is expected to support claims of health impacts, or one intended to prove a definite causal relationship between the intervention and the outcome (see Box 1.4).

#### Box 1.4. Timing and resources

1. At what stage of implementation is the evaluation meant to occur – beginning, during or at the end of implementation?
2. How much time is available to carry out evaluation activities?
3. What resources (human, financial, physical) are available to support evaluation activities?

## Step 7. Develop an M&E implementation plan

Once the study objectives, underlying framework and study design have been established, an implementation plan needs to be developed to provide a realistic roadmap of the timeline, resources and activities required to design and implement M&E activities. While there are various types of implementation plans (6–9), one common feature is a table that summarizes the basic activities, resources and time frame for the planned project. At a minimum the M&E implementation plan should include the following:

- **A structured list of activities and sub-activities:** Define and list the distinct activities and sub-activities that need to be carried out to implement each piece of the M&E framework. Examples of activities include the procurement of supplies, hiring and training of staff, development of M&E features in mobile applications, development of manuals or standard operating procedures (SOPs), collection of project or survey-based quantitative and qualitative data, establishment and implementation of mechanisms for data quality-assurance, data cleaning, analysis, interpretation, communication and dissemination.
- **Responsible persons assigned to activities:** Discuss the plan with all stakeholders and assign responsibility for various activities to specific staff members to ensure accountability. This list should include the name or job title of a focal point or person responsible for implementing each activity.
- **A timeline and target dates:** Specify a timeline for implementation, including dates when each activity should be carried out and/or the deadlines for completion of each activity. During implementation, this plan can be used as a tool to monitor fidelity of implementation activities to the implementation plan.
- **The budget and details of other resources required:** Plan the budget and required resources for each component of each project activity. If the activities are funded from multiple sources, the source of funding for each activity should be specified.

If the digital health intervention is supporting a broader health intervention, which is often the case, the implementation plan for M&E related specifically to the digital health system can be embedded within the M&E implementation plan for the larger health intervention or programme.

Table 1.2 provides an example of how some M&E activities may be delineated in an implementation plan. In this example, the digital health intervention uses SMS to deliver health information to pregnant women to support recommended visits to a health-care facility for antenatal care (ANC) and improve pregnancy outcomes. The project is interested in monitoring ANC visits and pregnancy outcomes in women participating in this intervention.

**Table 1.2. Illustrative format for an M&E implementation plan**

Activities	Sub-activities (responsible staff)	Timeline (Gant or due date)	Cost (source)
Objective 1: Monitor antenatal care service coverage and pregnancy outcomes.			
Activity 1: Develop a standard operating procedure (SOP) to collect data	Activity 1a: Convene a stakeholders meeting to decide on indicators and data to collect, and whether to collect it from the health information systems (health-care facility registries) or from pregnant women (project officer and M&E officer – ministry of health [MOH])	January 2016	\$XXX (MOH budget)
	Activity 1b: Draft the SOP (project officer)	February 2016	\$XXX (donor budget)
Activity 2: Train health workers and project staff on SOP	Activity 2a: Prepare training materials (project officer)	February 2016	\$XXX (donor budget)
	Activity 2b: Organize and conduct training (project officer)	March 2016	\$XXX (donor budget)

Table 1.2 gives an indication of what could be included in an implementation plan, but the level of detail supplied can be adapted to best serve the stakeholders' needs and to support effective monitoring of the activities in the plan as they are conducted and completed. The development of an M&E implementation plan promotes proactive calculation of data collection needs for the evaluation(s), allowing data to be collected prospectively, if needed. Data collected retroactively may suffer from biases that can affect the validity of the information.

## References

1. The mHealth planning guide: key considerations for integrating mobile technology into health programs. K4Health; 2015 (<https://www.k4health.org/toolkits/mhealth-planning-guide>, accessed 5 May 2016).
2. The MAPS Toolkit: mHealth Assessment and Planning for Scale. Geneva: World Health Organization; 2015 (<http://apps.who.int/iris/handle/10665/185238>, accessed 22 April 2016).
3. WHO evaluation practice handbook. Geneva: World Health Organization; 2013 ([http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf), accessed 25 April 2016).
4. Pritchett L, Samji S and Hammer J. It's all about MeE: using structured experiential learning ("e") to crawl the design space. Working Paper 322. Washington (DC): Center for Global Development; 2013 ([http://www.cgdev.org/sites/default/files/its-all-about-mee\\_1.pdf](http://www.cgdev.org/sites/default/files/its-all-about-mee_1.pdf), accessed 21 March 2016).
5. IFC Advisory Services Business Enabling Environment Business Line. The monitoring and evaluation handbook for business environment reform. Washington (DC): The World Bank Group; 2008 ([http://www.publicprivatedialogue.org/monitoring\\_and\\_evaluation/M&E%20Handbook%20July%2016%202008.pdf](http://www.publicprivatedialogue.org/monitoring_and_evaluation/M&E%20Handbook%20July%2016%202008.pdf), accessed 25 April 2016).
6. Handbook on planning, monitoring and evaluating for development results. New York (NY): United Nations Development Programme; 2009.
7. Umlaw F, Chitipo NN. State and use of monitoring and evaluation systems in national and provincial departments. *Afr Eval J*. 2015;3(1). doi:10.4102/aej.v3i1.134.
8. Project/programme monitoring and evaluation (M&E) guide. Geneva: International Federation of Red Cross and Red Crescent Societies; 2011 (<http://www.ifrc.org/Global/Publications/monitoring/IFRC-ME-Guide-8-2011.pdf>, accessed 21 March 2016).
9. The Project START monitoring and evaluation field guide. Atlanta (GA): United States Centers for Disease Control and Prevention (CDC); undated ([https://effectiveinterventions.cdc.gov/docs/default-source/project-start/Project\\_START\\_M\\_E\\_Field\\_Guide\\_10-1108.pdf?sfvrsn=0](https://effectiveinterventions.cdc.gov/docs/default-source/project-start/Project_START_M_E_Field_Guide_10-1108.pdf?sfvrsn=0), accessed 3 May 2016).





## Chapter 2: Setting the stage for monitoring and evaluation

**P**ositive results from M&E of digital health interventions are considered critical to support scale-up of the intervention since these results can lead to buy-in from stakeholders, such as donors and government entities. Hence, it is crucial that M&E objectives be aligned with overall project goals as well as expectations of stakeholders. Furthermore, developing an understanding of how project goals and activities relate to anticipated outcomes is necessary for selecting an appropriate study design and meaningful indicators of success.

This chapter lays the foundation for well aligned and well designed M&E efforts by elaborating on the fundamental questions of:

- What is the goal of your M&E efforts?
- How will you organize the process to achieve your M&E goals? and
- How will you measure the achievement of your M&E goals?

Part 2a introduces the process for articulating the anticipated benefits of the digital health intervention, using what are called **claims**, in an effort to align M&E efforts to stakeholder expectations, and to drive adoption and scale-up of the digital health intervention. Part 2b describes the process for developing an M&E framework to outline the process and rationale that helps to arrive at M&E research goals. Finally, Part 2c discusses the use of indicators and presents a generic listing of indicators judged to be useful for M&E of digital health interventions.

## Part 2a: Articulating claims



### HOW WILL THIS SECTION HELP ME?

This section will:

- ✓ Help you to articulate the “claims” of the digital health intervention that would serve as the basis for determining the M&E objectives and for defining the evidence needs for stakeholders.
- ✓ Provide illustrative evidence claim statements to guide the formulation of M&E objectives, hypotheses and indicators.
- ✓ Describe a step-wise approach to ensure that the claims are appropriate, measureable and of importance to identified stakeholders.

At the core of every digital health system or intervention is a **value proposition** – a statement describing the benefits to end-users, with an implicit comparator, which can be a non-digital intervention or an alternative digital product (1). Well crafted value propositions can drive the successful adoption and sustainability of digital health systems by persuasively communicating their value to the end-users (1, 2). For example, Dimagi states the value proposition for its CommCare platform as “Build mobile apps in days, not months”, indicating the speed and ease with which new projects can be customized and deployed using the platform (3). Value propositions describe (i) which end-user needs are met by the digital health system and how, (ii) why the digital health system is innovative, and (iii) why the digital health system is superior to the standard of care or status quo (1). Value propositions are important precursors to the development of a business model describing the project’s goals and plans for scaling up and achieving sustainability (2). Value propositions are based on a verified end-user need (e.g. through formative evaluation; see Chapter 4, Part 4a) and a validated digital health system (e.g. through monitoring and/or summative evaluation; see Chapter 3, and Chapter 4, Parts 4a and 4b) (1). Claims about the digital health intervention are based on assumptions about end-user needs and/or



**Claim:** A statement of anticipated benefits of the digital health system or intervention.

**Value proposition:**  
A statement describing the benefits to end-users, with an implicit comparator, which can be a non-digital intervention or an alternative digital product (1).

the effectiveness of the digital health system. Articulating intended or expected future claims can help to define the M&E objectives, and this is one of the first steps in crafting the project's value proposition.

In order to convince stakeholders that the digital health intervention is suitable for scale-up, project managers must craft one or more value proposition statements related to the intervention's efficacy, effectiveness or cost-effectiveness. Using a claims-based approach to inform M&E objectives offers several advantages. First, articulating claims early on can help align M&E efforts to stakeholder expectations. This ensures that projects are generating an evidence base for the components of the digital health intervention that are of greatest value to stakeholders, and this may in turn spur stakeholder investments or adoption of the product. Second, articulating claims allows project managers to identify the key processes and outcomes that need to be monitored or evaluated. Doing so can potentially reduce costs by streamlining M&E efforts to focus on the most critical pieces of evidence needed to support scale-up. Finally, claim statements can guide the choice of indicators that can best be used to measure key processes and outcomes. All project claims must eventually be articulated as measurable M&E objectives. Box 2.1 illustrates the differences between claim statements, M&E objectives, hypotheses and indicators. Part 2c describes the process of incorporating claims into M&E efforts, articulating them as measurable objectives and using them to guide the selection of indicators.

#### **Box 2.1. Illustrative examples of a claim, M&E objective, hypothesis and indicator**

**Claim:** Proactive SMS/text message vaccination reminders to mothers improve coverage of measles vaccine in their children.

**Evaluation objective:** Measure change in measles vaccine coverage among children aged 12–23 months whose mothers receive text message reminders on upcoming vaccinations compared to those whose mothers receive routine immunization services with no text message reminders.

**Hypothesis:** Text message reminders on upcoming vaccinations to mothers improve coverage of measles vaccine among their children aged 12–23 months by 15% compared to no text message reminders after one year of implementation.

**Indicator:** Percentage of children aged 12–23 months receiving measles vaccination through routine immunization services in the preceding year.

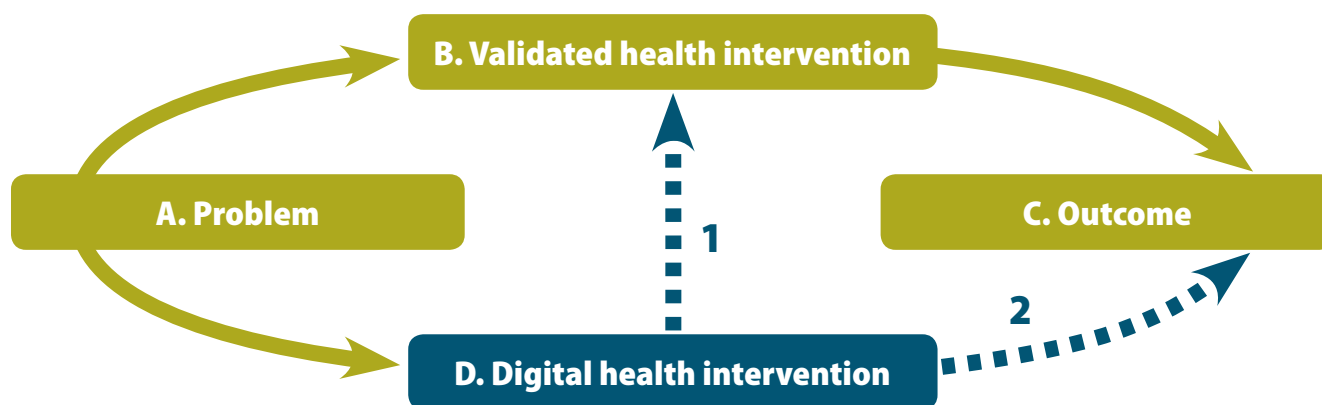
## **A claims-based approach to defining M&E objectives for digital health interventions**

Developing an evaluation strategy and appropriate claims involves determining whether the digital health system being considered is merely a means of improving the quality or coverage of an intervention known to be effective, or whether it instead constitutes a novel intervention in itself, the effectiveness of which is, as yet, unknown. If it is the former type, then, given the costs involved, there may not be a great need to gather further evidence of the efficacy or effectiveness of a health intervention before we recommend the use of a digital health system to improve the quality and coverage of that intervention. For example, digital health systems may be used to optimize the delivery of vaccines in terms of timing, coverage and completeness of the vaccination schedule, while the vaccines themselves have already been previously tested for efficacy (i.e. they have been shown to reduce rates of infection or illness in prior studies) and administered through other programmes.

However, using a digital enhanced algorithm to improve clinical decision-making may be different in nature to the example of using such approaches to improve vaccine coverage. In the case of this electronic decision support system, the number of variables for which the efficacy is “unknown” increases considerably: the algorithm itself, the mode of delivery, and the use of a digital application to support care provision.

Claims made in relation to most digital health interventions fall into one of two pathways to effect a health outcome (see Figure 2.1) (4).

**Figure 2.1. Defining the role of a digital health intervention**



**Pathway 1:** Are you evaluating the added benefit of the digital health system to optimize the delivery of an existing or already validated health intervention, and thereby improve health outcomes? For example: Do digital supply chain systems improve the coverage of childhood vaccines?

If you answer “yes” to this Pathway 1 question, you are working with a digital health system that has a well established underlying evidence base; it delivers an intervention with a proven health impact. The beneficial health impact of the intervention has been established through prior research. Therefore, evaluation efforts should focus on outcome measures (changes in health status, disease prevalence, etc.) and/or process/functionality measures (numbers of persons trained, etc.). In this case, claims for the digital health system can focus on the performance of the digital health intervention’s delivery system, which provides added benefit or comparative effectiveness, such as improving coverage of the intervention (e.g. of childhood vaccines), which will have a positive impact on population health.

An example of a claim statement when following Pathway 1 is: Digital health intervention X will result in an increase in coverage in children under the age of 1 year of measles vaccinations administered through the routine immunization programme.

**Pathway 2:** Are you evaluating the effectiveness of the digital health intervention to directly and independently trigger a health outcome (i.e. where the effectiveness is not yet known)? For example: Do electronic decision support systems improve the quality of services provided by health-care providers?

If you answer “yes” to the Pathway 2 question instead, you are working with a digital health system that is deemed to be a novel intervention in and of itself, where there is not a strong underlying evidence base for the intervention. In this case, validation of the approach and evaluation of the health impact should be considered before claims can be formulated. For projects that focus on the use of new interventions, claims may relate to the efficacy, effectiveness or cost–effectiveness of the intervention, including any anticipated impacts on health or behaviour. M&E efforts for these projects may capture **process, outcome and impact** measures.

An example of a claim statement when following Pathway 2 is: Digital health intervention X will improve clinical decision-making among health-care providers, through new algorithms that inform electronic decision support.

There may be scenarios in which the purpose of the evaluation is to answer both questions; to follow both pathways. In either scenario, the question for decision-makers and stakeholders is, What claim do we want to make about the digital health intervention? In other words, the M&E team must decide if they want to suggest that it (i) enhances delivery of services with known efficacy/effectiveness or (ii) has a potentially independent and untested effect on a given outcome of interest.

## Linking claims with the Sustainable Development Goals

With the adoption of the 2030 Agenda for Sustainable Development in September 2015, and the 17 Sustainable Development Goals (SDGs)<sup>3</sup>, health systems and stakeholders are interested in innovative approaches for achieving universal health coverage (UHC) objectives. In this context, it may be useful to structure claims for digital health interventions, especially those focusing on existing evidence-based health interventions, on the determinant layers of UHC (see Table 2.1) (5).

**Table 2.1. Illustrative claim statements based on determinant layers of universal health coverage (UHC)**

Determinant layers of UHC		Illustrative digital health strategies to close performance gaps	Illustrative claim statements
<b>Accountability</b>	Accountability coverage	<ul style="list-style-type: none"> <li>Registries and vital events tracking</li> <li>Electronic medical records</li> <li>Data collection and reporting</li> </ul>	Digital health intervention X will facilitate electronic birth registration of newborns.
	Supply		
	Availability of commodities and equipment	<ul style="list-style-type: none"> <li>Supply chain management</li> <li>Counterfeit prevention</li> </ul>	Digital health intervention X will reduce stock-outs of drug Y in N districts.
	Availability of human resources	<ul style="list-style-type: none"> <li>Human resource management</li> <li>Provider training</li> <li>Telemedicine</li> </ul>	Digital health intervention X will increase the availability of providers trained in identifying signs of postpartum haemorrhage in new mothers through provision of multimedia education content.
	Availability of health-care facilities	<ul style="list-style-type: none"> <li>Hotlines</li> <li>Client mobile applications</li> <li>Client information content subscriptions</li> </ul>	Digital health intervention X will provide information to clients about family planning methods on demand.
<b>Demand</b>	Contact coverage	<ul style="list-style-type: none"> <li>Behaviour change communication</li> <li>Incentives</li> </ul>	Digital health intervention X will provide phone consultations with health-care providers to clients on demand.
	Continuous coverage	<ul style="list-style-type: none"> <li>Persistent electronic health records</li> <li>Provider-to-provider communication</li> <li>Work planning</li> <li>Reminders</li> </ul>	Digital health intervention X will alert community-based vaccinators about children who are overdue for routine immunization services.
<b>Quality</b>	Effective coverage	<ul style="list-style-type: none"> <li>Decision support</li> <li>Point-of-care (POC) diagnostics</li> <li>Telemedicine</li> <li>Reminders</li> <li>Incentives</li> </ul>	Digital health intervention X will improve community health workers' adherence to clinical protocols.
<b>Cost</b>	Financial coverage	<ul style="list-style-type: none"> <li>Mobile financial transactions</li> </ul>	Digital health intervention X will use mobile money vouchers to subsidize travel costs associated with facility-based deliveries for pregnant women.

Source: adapted from Mehl and Labrique, 2014 (5).

<sup>3</sup> Further information available at: <https://sustainabledevelopment.un.org/sdgs>

## Steps in a claims-based approach

The key steps in a claims-based approach are described below and in Figure 2.2.

### i. Map stakeholders

Stakeholders are defined as entities (individuals or organizations) that have a vested interest in the digital health system or intervention, in the capacity of being a decision-maker, project staff or end-user (6). Members of the scientific or digital health communities may also be considered stakeholders of digital health systems or interventions. The latter may have direct or indirect interests in the products, strategies, data generated or policies influenced by the digital health intervention.

The claims-based approach begins with the identification and listing of key stakeholders associated with the project. Projects using digital health technologies are typically multidisciplinary and engage a wide range of stakeholders, each of whom may contribute different resources to the projects and, hence, have different expectations for returns from the projects. The selected stakeholders could represent existing partnerships, those being pursued and/or those considered important for scale-up. The latter category of stakeholders is especially important as they can potentially determine whether the digital health intervention will be successfully scaled up or not. Therefore, identifying and including these stakeholders early on in the evidence-generation process can ensure that a project has sufficient and relevant data to achieve buy-in when poised for scale-up. Managers of a digital health project may choose to embark on a formal stakeholder mapping exercise to identify and prioritize relevant stakeholders (6, 7). The goal of such an exercise is usually to narrow down the list of stakeholders to those who may be “primary” or “key” stakeholders (6).

**Figure 2.2. Steps in the evidence-claims approach for informing M&E objectives**

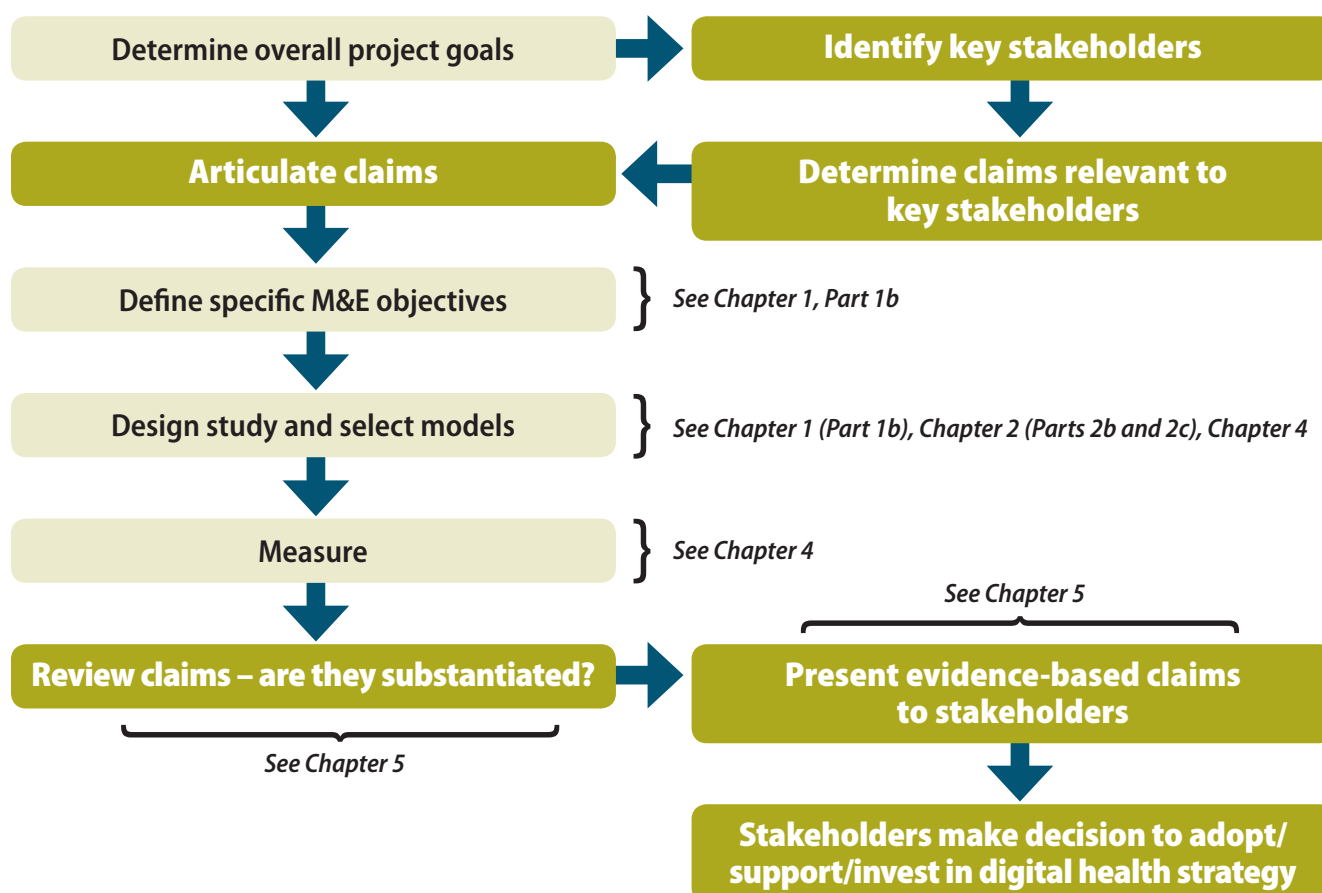




Table 2.2 lists the categories of stakeholders who may be associated with a digital health project, their roles and example categories of claim statements that may be relevant to them.

## ii. Clarify expectations

A common pitfall in the synthesis of claim statements is the assumption that one knows what the stakeholders expect (or worse, the assumption that the stakeholder has no expectations at all). Clarifying stakeholder expectations early can ensure that the claim statements are relevant and focused, potentially preventing allocation of resources to low-priority processes and outcomes. Ways of engaging with stakeholders may range from reviewing their annual reports or strategic plans to learn about their interests and priorities, to active networking, and proposing or initiating active collaboration (see Box 2.2) (8). Focus group discussions or in-depth interviews with key informants (see Part 4a) may be ways to gather information on stakeholder perceptions, needs and expectations.

### Box 2.2. Examples of questions for stakeholders

- What are the top three priorities for your organization?
- What are your key expectations from this project?
- In what ways can this project add value to your organization's mission?
- What is the main outcome you expect this digital health intervention to achieve?

## iii. Articulate claims

The expectations outlined by the stakeholders can then be articulated in the form of claim statements. When articulating claims:

- Begin by listing all the claims you can think of that are relevant for each stakeholder. Then narrow down the claims to the top three claims you think are most important for each stakeholder.
- Avoid vague statements. “High profile innovation improves coverage of President’s Safe Motherhood Programme” is better than stating “Innovative project adopted by government”.
- Claim statements may specify processes, outcomes or health impact.
- Claims should ultimately be measurable. To achieve this, claims may be articulated in the form of specific M&E objectives or indicators (Part 2c).

## iv. Measure claims

Claims may be measured or substantiated during M&E activities (see Chapter 2, Part 2b, and Chapters 3 and 4). Claims may also be substantiated through process documentation, training, content development, fundraising or creation of the technologies/products.

## v. Update claims to match evidence base

Once the claim statements have been measured, it is important to revise the claims to match the evidence, if needed. For instance, if only a 10% increase in measles vaccination coverage was seen rather than the anticipated 15% increase, then your claim statement must be revised to reflect that statistic. A data-mapping exercise to assist with this process is presented in Chapter 5.



**Table 2.2. Illustrative categories of digital health stakeholders and claims**

Stakeholder category	Role in supporting digital health intervention	Illustrative claim categories relevant to the stakeholder	Illustrative claim statements
<b>Government entities</b> Example: Ministry of Health	Implementation partners and/or target adopters of the digital health interventions	<ul style="list-style-type: none"> <li>Alignment with country processes and governance needs (e.g. for policy-making)</li> <li>Improved health system functioning (e.g. better performance of health workforce, availability of commodities, coverage of services)</li> <li>Improved RMNCAH impact (e.g. lower maternal and infant morbidity and mortality)</li> </ul>	<p><b>Claim 1:</b> Proactive text message vaccination reminders to mothers improve coverage of measles vaccine in their children.</p> <p><b>Claim 2:</b> Mobile-phone-assisted electronic birth registration is a cost-effective way to enumerate newborns.</p>
<b>Private sector organizations</b> including mobile network operators Example: Vodafone	Mobile network operator (MNO) partners	<ul style="list-style-type: none"> <li>Increased use of network (e.g. compared to competitors)</li> <li>Cost-effective solution (e.g. low per capita cost)</li> <li>Adequate infrastructure for maintenance and scale-up (e.g. ability of operator to support future efforts)</li> </ul>	<p><b>Claim 1:</b> Providing maternal health information to the customer base may serve to reduce churn and promote brand loyalty.</p> <p><b>Claim 2:</b> Digital health supply chain management system allows distribution tracking and verification of authenticity for drugs from factory to consumers.</p>
<b>Donors</b> Example: Bill & Melinda Gates Foundation	Funders	<ul style="list-style-type: none"> <li>High impact</li> <li>Alignment with strategic plan of donor</li> <li>High return on investment (e.g. improved health-care delivery to disadvantaged populations through use of technology)</li> </ul>	<p><b>Claim 1:</b> Digital health data collection platform is used by 6000 community health workers in 80 districts providing maternal and child health services.</p> <p><b>Claim 2:</b> 200 000 women have been screened for cervical cancer using mobile-phone-assisted digital cervicography.</p>
<b>Technical agencies</b> Example: World Health Organization	Technical support and guidance related to health domain	<ul style="list-style-type: none"> <li>Improved reporting and monitoring systems</li> <li>Interoperable systems</li> <li>Transferability to other countries or health systems</li> <li>Sustainable integration into health system</li> </ul>	<p><b>Claim 1:</b> Mobile-phone-based interactive voice response (IVR) system is a feasible way to facilitate timely routine surveillance of dengue fever outbreaks by community health workers.</p> <p><b>Claim 2:</b> The mobile data collection platform is interoperable with a popular health management information system (HMIS) deployed in over 50 countries.</p>
<b>Beneficiaries or clients</b> Example: Pregnant women; community health workers	Target audience	<ul style="list-style-type: none"> <li>Access to quality, equitable health care</li> <li>Access to affordable health care</li> <li>Improved RMNCAH impact (e.g. low maternal and infant morbidity and mortality)</li> </ul>	<p><b>Claim 1:</b> Gestational-age-specific health information is delivered through accessible, low-cost mobile phone channels to support healthy pregnancy.</p> <p><b>Claim 2:</b> The digital health system facilitates management of chronic disease through daily tracking of diet, exercise and medications.</p>
<b>Nongovernmental organizations</b> Example: Médecins Sans Frontières	Implementation partners	<ul style="list-style-type: none"> <li>Context-specific solution</li> <li>Adequate support for quality assurance, training and maintenance</li> <li>Local capacity-building</li> </ul>	<p><b>Claim 1:</b> Training needs are low for implementation of the digital health system.</p> <p><b>Claim 2:</b> The digital health system is highly stable with low error rates in data transmission and, hence, low maintenance needs.</p>

Stakeholder category	Role in supporting digital health intervention	Illustrative claim categories relevant to the stakeholder	Illustrative claim statements
<b>Digital health community</b> including any local or national Technical Working Groups (TWGs)  Example: Tanzania mHealth community of practice	Peers, future adopters of interventions/ technologies developed by the project	<ul style="list-style-type: none"> <li>■ High motivation for use and desirability</li> <li>■ High stability and low cost of technology</li> <li>■ Improved RMNCAH impact (e.g. low maternal and infant morbidity and mortality)</li> </ul>	<b>Claim 1:</b> Health information content for promoting smoking cessation has been adapted and validated for text messaging.  <b>Claim 2:</b> There is high satisfaction and acceptability of the mobile network closed user group for health providers.
<b>Scientific community</b>  Example: Global Symposium on Health Systems Research Conference	Peers, future adopters of interventions/ technologies developed by the project	<ul style="list-style-type: none"> <li>■ Based on validated clinical guidelines</li> <li>■ Rigorous methodology used for evaluation</li> <li>■ Significant RMNCAH impact (e.g. low maternal and infant morbidity and mortality)</li> </ul>	<b>Claim 1:</b> Text message reminders are a cost-effective strategy to improve HIV treatment adherence.  <b>Claim 2:</b> Patients with diabetes using the mobile-phone-based diabetes management application show a reduction in Haemoglobin A1c levels compared with those not using the application in a randomized controlled trial.

RMNCAH: reproductive, maternal, newborn, child, and adolescent health

## Part 2b: Developing an M&E framework



### HOW WILL THIS SECTION HELP ME?

This section will:

- ✓ Describe the variety of established frameworks that are relevant for the M&E of digital health interventions.
- ✓ Demonstrate the appropriate use of different frameworks according to identified M&E needs and objectives.
- ✓ Highlight real examples of various M&E frameworks as applied to digital health projects.

In Part 1b, Step 2, we highlighted the importance of developing an underlying conceptual framework to help you to define and understand your project goals and objectives and to conceptualize the relationship between these. Conceptual frameworks are also used to define the underpinning project activities required to achieve your goals and objectives, and to describe the anticipated outcomes. In Table 2.3 we outline some of the most commonly used frameworks: (i) **conceptual framework**; (ii) **results framework**; (iii) **logical framework**; and (iv) **theory of change**. This section provides a synthesis of these frameworks and illustrates the application of frameworks to the M&E of digital health interventions.



**Conceptual framework (also known as theoretical or causal framework):** A diagram that identifies and illustrates the relationships among factors (systemic, organizational, individual or other) that may influence the operation of an intervention and the successful achievement of the intervention's goals (9). The purpose is to facilitate the design of the digital health intervention or project and provide a theoretical basis for the approach.

**Results framework:** A "graphic representation of a strategy to achieve a specific objective that is grounded in cause-and-effect logic" (10). The main purpose of this type of framework is to clarify the causal relationships that connect the incremental achievement of results to intervention impact.

**Logical framework/logic model:** A management and measurement tool that summarizes what a project intends to do and how, what the key assumptions are, and how outputs and outcomes will be monitored and evaluated. The aim of a logic model is to clarify programme objectives and aid in the identification of expected causal links between inputs, processes, outputs, outcomes and impacts (11).

**Theory of change:** A theory of change is a causal model that links outcomes and activities to explain how and why the desired change is anticipated to occur (12). Theory-based conceptual frameworks are similar to logic models but aim to provide a greater understanding of the complex relationship between programme activities and anticipated results.

**Inputs:** The financial, human, material and intellectual resources used to develop and implement an intervention. In this Guide, inputs encompass all resources that go into a digital health intervention.

**Processes:** The activities undertaken in the delivery of an intervention – a digital health intervention for the purposes of this Guide.

**Outputs:** The direct products/deliverables of process activities in an intervention (13). From a digital health perspective, outputs can include improvements in performance and user adoption.

**Outcomes:** The intermediate changes that emerge as a result of inputs and processes. Within digital health, these may be considered according to three levels: health systems, provider and client.

**Impact:** The medium- to long-term effects produced by an intervention; these effects can be positive and negative, intended and unintended (14).



Knowledge for Health (K4Health) describes the use of different types of frameworks (see Figure 2.3). Further information and background on the general use of frameworks is available on the K4Health website (15).

**Figure 2.3. Toolkits by K4Health webpage on frameworks**



Source: K4Health, 2016 (15).

## Conceptual framework

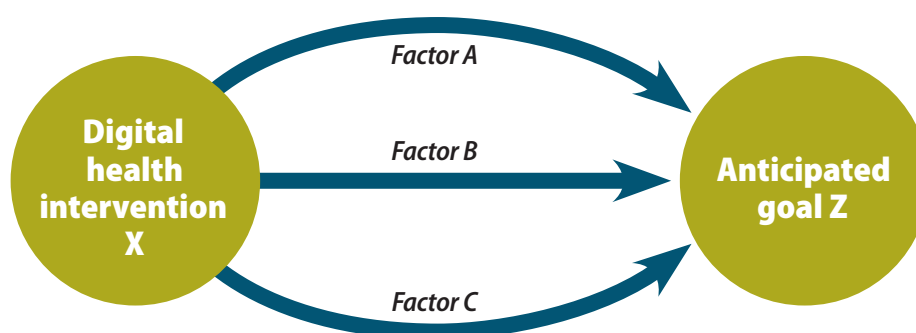
Conceptual frameworks, also known as theoretical or causal frameworks, are diagrams that identify and illustrate the relationships among factors (systemic, organizational, individual or other) that may influence the operation of an intervention and the successful achievement of the intervention's goals (9). They aim to facilitate the design of your digital health intervention or project and provide a theoretical basis for your approach. As described by Earp and Ennett (1991), a conceptual model is a visual “diagram of proposed causal linkages among a set of concepts believed to be related to a particular public health problem” (16). “Concepts” are represented by boxes and include all salient factors that may influence programme/project operation and successful achievement of the goals. Processes are delineated by arrows, which are intended to imply causality (see Figure 2.4).

To create a conceptual framework:

- start with your digital health intervention [X];
- define your “endpoint” or the anticipated goal [Z];
- identify the pathway (including intermediate “goal posts” A, B, C, etc.) that connects your intervention with the desired goal (based on evidence available).

As a rule of thumb, only include factors that can be operationally defined and measured. Then working from left to right, and using arrows to imply causality, connect the factors, which in series are anticipated to yield your desired goal. In other words, your framework charts your hypothesis that intervention X can cause goal Z, by first changing factor A, then factor B, then factor C, etc.

**Figure 2.4. Template diagram for a conceptual model or framework**



**Table 2.3. Frameworks for defining the scope of M&E activities**

Type	Description	Purpose
<b>Conceptual framework</b>	<ul style="list-style-type: none"> <li>■ A diagram that identifies and illustrates the relationships among factors (systemic, organizational, individual or other) that may influence programme/project operation and the successful achievement of programme or project goal(s)</li> </ul>	<ul style="list-style-type: none"> <li>■ Identifies factors that influence programme goals (e.g. service utilization) in order to highlight enablers and barriers in the pathway</li> <li>■ Provides a perspective for understanding programme objectives within the context of factors in the operating environment</li> <li>■ Clarifies analytical assumptions and their implications for programme possibilities or limitations on success</li> </ul>
<b>Results framework</b>	<ul style="list-style-type: none"> <li>■ A planning and management tool</li> <li>■ A diagram that identifies and illustrates causal relationships between programme objectives and observed impact</li> <li>■ Links the outcome with hypothesis or theory about how desired change (impact) is anticipated to occur through lower- and higher-level objectives and immediate and lower-level results</li> </ul>	<ul style="list-style-type: none"> <li>■ Allows managers to gauge progress towards the achievement of results and to adjust programme activities accordingly</li> <li>■ Provides clarified focus on the causal relationships that connect incremental achievement of results to the programme impact</li> <li>■ Clarifies project/programme mechanics and relationships between factors that suggest ways and means of objectively measuring the achievement of desired impact</li> </ul>
<b>Theory of change</b>	<ul style="list-style-type: none"> <li>■ Both a process and product</li> <li>■ A causal model that links outcomes and activities to explain how and why the desired change is anticipated to occur (12)</li> <li>■ Explains “how you see the world, and how change happens, and how you are going to intervene based on that understanding” (12)</li> </ul>	<ul style="list-style-type: none"> <li>■ Describes the sequence of events that is expected to yield a desired outcome</li> <li>■ Provides an integrated approach for designing, implementing and evaluating programme activities</li> <li>■ Describes how and why you think change will occur through a flexible diagram showing all pathways that may lead to change</li> </ul>
<b>Logical framework</b>	<ul style="list-style-type: none"> <li>■ A management and measurement tool</li> <li>■ A model that summarizes what a project intends to do and how, what the key assumptions are, and how outputs and outcomes will be monitored and evaluated</li> <li>■ Diagrams that identify and illustrate the linear relationships flowing from programme inputs, through processes and outputs, to outcomes</li> </ul>	<ul style="list-style-type: none"> <li>■ Provides a streamlined interpretation of planned use of resources and goals</li> <li>■ Clarifies project/programme assumptions about linear relationships between key factors relevant to intended goals</li> <li>■ Provides a way of measuring success and making resource allocation decisions</li> </ul>

## Results framework

A results framework, as described by USAID (2010), is a “graphic representation of a strategy to achieve a specific objective that is grounded in cause-and-effect logic” (10). The main purpose of this type of framework is to clarify the causal relationships that connect the incremental achievement of results to programme impact. The process of developing a results framework helps to:

- build consensus and ownership for the activities that comprise the programme
- identify ways to measure the achievement of desired programme goals
- select appropriate inputs needed to achieve objectives
- establish the foundation for designing M&E plans and
- refine the definition of programme objectives.

A results framework includes a hypothesis or theory about how the desired change is anticipated to occur. This includes linkages between lower- and higher-level objectives and, ultimately, the resulting outcome. Following are the steps to create a results framework.

1. **Develop a hypothesis** of your intervention’s anticipated effect in yielding an outcome of interest.
2. **Finalize programme objectives** that balance ambition and accountability, and which also take into account programme history, the magnitude of the development problem, time frame and availability of resources (10).
3. **Identify intermediate results** that are measurable.

4. **Review the intermediate results** to confirm the logic and ensure that their achievement will lead to the next higher-level objective.
5. **Identify critical assumptions.**
6. **Identify preliminary performance measures**, drawing from baseline data, which specify measurable and attainable targets (10).

Box 2.3 (including Figure 2.5) describes the Mobile Technology for Community Health (MOTech) Initiative in Ghana. Figure 2.6 provides an illustrative example of a results framework developed for the MOTech programme. The results framework illustrates the relationships between the desired programme goal (improved maternal and child health) and the immediate results and lower-level results that were anticipated to facilitate achievement of the goal. Immediate results (IR) included:

1. improved coverage of the Mobile Midwife application and access to health information
2. improved maternal and child health behaviour and knowledge
3. improved management of child health data at the district level
4. improved ownership of MOTech by the Ghana Health Service and
5. demonstrated sustainability of MOTech.

Lower-level results (LLR) required to achieve these, as well as illustrative indicators, are presented in boxes branching off from the IR in Figure 2.6.

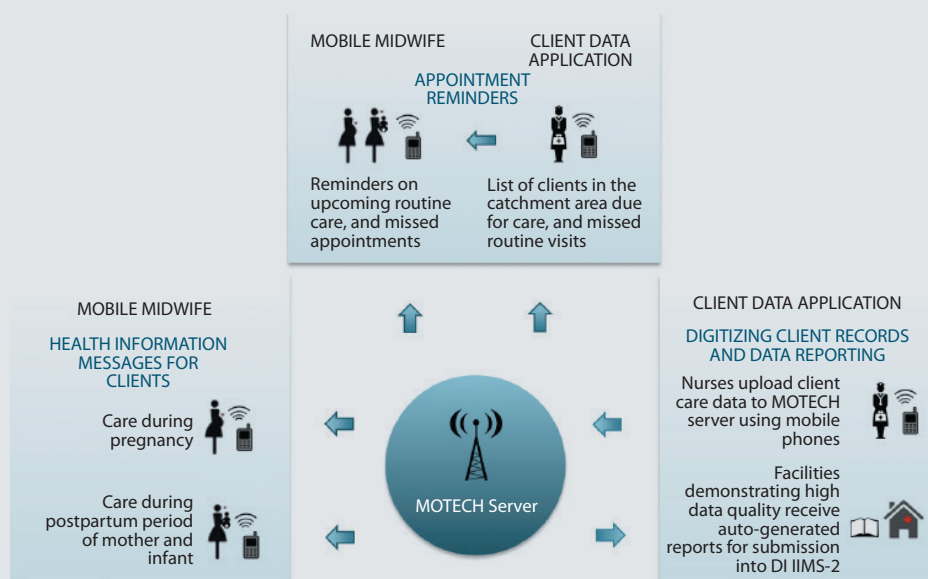
### Box 2.3. Description of the MOTech Project in Ghana

Grameen Foundation worked with the Ghana Health Service from 2009 to 2014 to develop and implement the MOTech (Mobile Technology for Community Health) platform, which delivers two interrelated mobile applications in Ghana – Mobile Midwife and the Client Data Application – to address some of the information-based drivers of maternal, newborn and child health in Ghana.

**Mobile Midwife** delivers pre-recorded voice messages to women, providing stage-specific educational information about pregnancy and infant health for them in their own languages.

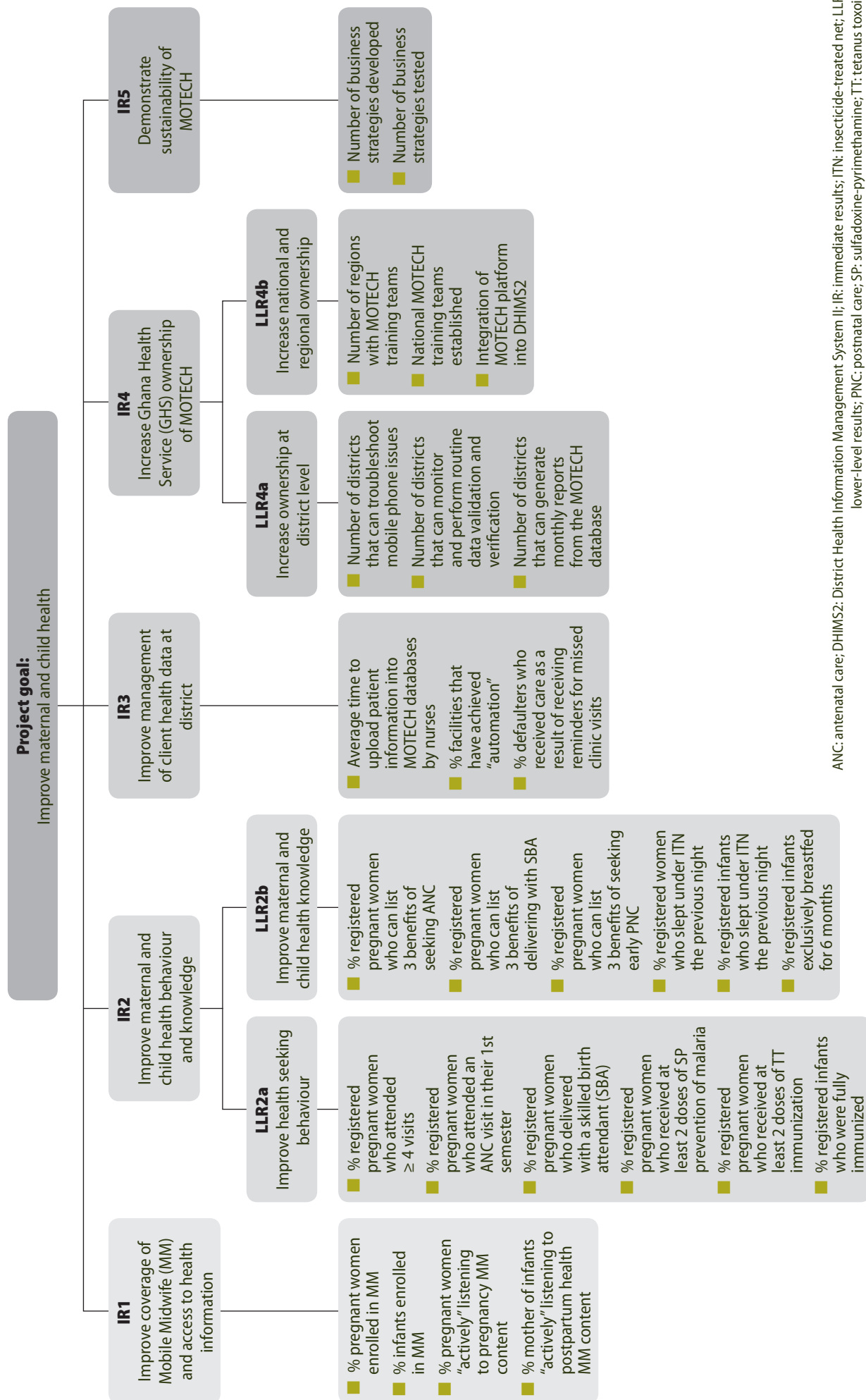
**Client Data Application** enables community health nurses based at front-line health-care facilities to use a mobile phone to electronically record the care given to patients, which facilitates monthly reporting and makes it easier for them to identify women and infants in their area who are due or overdue for care.

**Figure 2.5. MOTech's Mobile Midwife and Client Data Application**



Source: MOTech, unpublished data, 2016. For further information, see Grameen Foundation, 2015 (17).

**Figure 2.6. Illustrative results framework for MOTECH in Ghana**



ANC: antenatal care; DHIMS2: District Health Information Management System II; IR: immediate results; ITN: insecticide-treated net; LLR: lower-level results; PNC: postnatal care; SP: sulfadoxine-pyrimethamine; TT: tetanus toxoid

Source: MOTECH, unpublished data, 2016. For further information, see Grameen Foundation, 2015 (17).



## Logical framework

A logical framework is a management and measurement tool that summarizes what a project intends to do and how, what the key assumptions are, and how outputs and outcomes will be monitored and evaluated. The aim of a logical framework is to clarify programme objectives and aid in the identification of expected causal links between inputs, processes, outputs, outcomes and impacts (11). A logical framework is created to provide a graphical representation that can serve as a catalyst for engaging and communicating with key stakeholders, including implementers, in an iterative process, often in the wake of changes in programme design or implementation. Figure 2.7 provides an illustrative logical framework for the MomConnect initiative in South Africa (MomConnect is described in Box 2.4, and Figure 2.8).

Logical frameworks link inputs (programme resources) with processes (activities undertaken in the delivery of services), outputs (products of processes), outcomes (intermediate changes) and impacts.

**Inputs** are defined as the financial, human, material and intellectual resources used to develop and implement an intervention. In this Guide, inputs encompass all resources that go into a digital health intervention. In this model, technology inputs (e.g. software application development) are differentiated from programmatic inputs aimed at providing health services. Programmatic inputs (human resources, training, and development of other materials) are distinguished from policy inputs, which relate to linkages with treatment and care as well as issues such as affordability, including user fees.

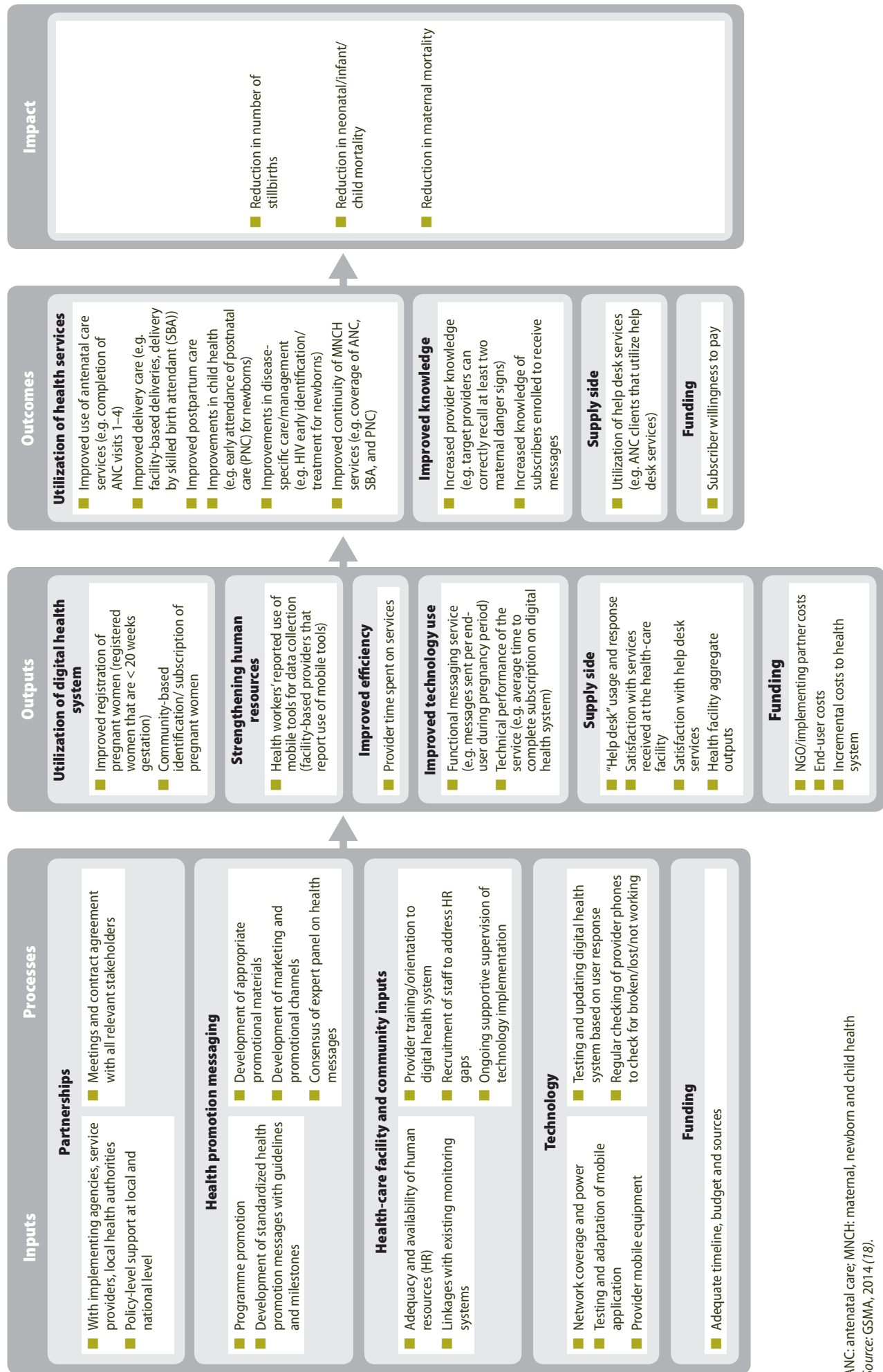
**Processes** are defined as the activities undertaken in the delivery of an intervention – a digital health intervention for the purposes of this Guide. Processes may include training courses and partnership meetings, as well as the activities required to test and update the digital health system based on user response. For digital health interventions that are in the latter stages of maturity and evaluation (e.g. effectiveness to implementation science), beyond initial inputs to the recruitment and training of providers, programmes will need to monitor supportive supervision, provider performance, attrition and training courses (refresher and initial) provided during implementation.

**Outputs** are defined as the direct products/deliverables of process activities in an intervention (13). From a technological perspective, technology inputs (e.g. hardware/devices and software) coupled with the capacity-building to ensure their appropriate and sustained use, correspond to changes in programme outputs – including improvements in performance and user adoption. Ultimately these technological outputs are anticipated to correspond to improved functioning of health systems (governance, human resources, commodity management) and service delivery. Improvements in service delivery include increased outreach and follow-up (increased number of provider visits); improved availability and quality of services; improved service integration; and increased proficiency and accountability among health-care providers.

**Outcomes** refer to the intermediate changes that emerge as a result of inputs and processes. Outcomes can be assessed at three levels: health systems, provider and client. At the health systems level, outcomes encompass domains of efficiency (technical and productive), increased service responsiveness to meet client needs, and increased coverage of target health services. At the provider level, increases in knowledge, productive efficiency (e.g. time allocation), and quality of care can be anticipated as outcomes. Finally, at the client level, digital health interventions can be expected to bring outcomes including changes in knowledge, efficiency (technical and productive), service responsiveness, adherence to treatment protocol and, ultimately, demand for services.

**Impact:** The impact of health interventions can be defined as the medium- to long-term effects produced by an intervention; these effects can be positive and negative, intended and unintended (14). For digital health interventions that aim to improve the delivery of health interventions with known efficacy/effectiveness, generating data on health impact may not be needed (as in Pathway 1 from Part 2a). For digital health interventions that are novel interventions in themselves, with limited evidence of effectiveness, gathering such evidence first may be essential to inform decision-making on the appropriate allocation of resources for the intervention (as in Pathway 2 from Part 2a). Accordingly, health impact may be considered according to domains of health systems performance (e.g. increased provider time spent on clinical care), population health (e.g. reductions in morbidity and mortality), as well as additional population benefits, including reductions in catastrophic health-care expenditures for households.

**Figure 2.7. Illustrative logic model for MomConnect in South Africa**



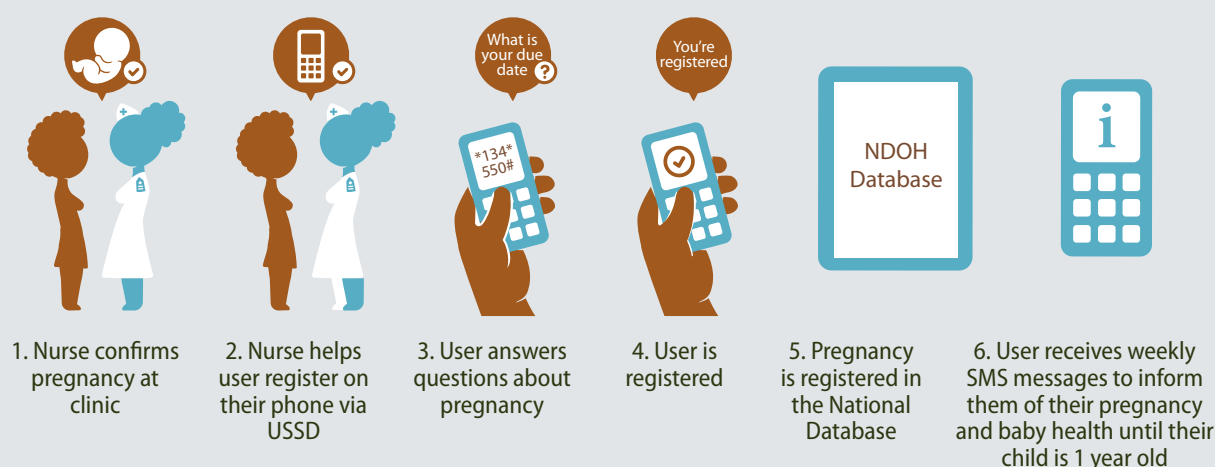
ANC: antenatal care; MNCH: maternal, newborn and child health  
Source: GSMA, 2014 (18).

### Box 2.4. Description of MomConnect in South Africa

MomConnect is a South African National Department of Health initiative to use mobile phone SMS technology to register every pregnant woman in South Africa. MomConnect aims to strengthen demand for and accountability of maternal and child health services in order to improve access, coverage and quality of care for mothers and their children in the community. Further information is available at: <http://www.rmchsa.org/momconnect/>

Once registered, each mother receives stage-based SMS messages to support her own and her baby's health. Since August 2014, more than 500 000 pregnant women have been registered throughout South Africa. Phase II aims to expand services at the community level through community health workers.

**Figure 2.8. MomConnect illustrated: how it works**



Source: South Africa NDOH, 2015 (19).

## Theory of change

A theory of change is a causal model that links outcomes and activities to explain how and why the desired change is anticipated to occur (12). Theory-based conceptual frameworks are similar to logic models but aim to provide a greater understanding of the complex relationship between programme activities and anticipated results. Most notably, they do not assume a linear cause-and-effect relationship (11), but rather encourage the mapping of multiple determinants or causal factors as well as underlying assumptions, which can be tested and measured.

To test a theory of change you need to consider the following questions:

- What is the target population your digital health intervention aims to influence or benefit?
- What results are you seeking to achieve?
- What is the expected time period for achieving the anticipated results?
- What are the activities, strategies and resources (human, financial, physical) required to achieve the proposed objectives?
- What is the context (social, political and environmental conditions) in which you will work?
- What assumptions have you made?

Theories of change should minimally contain the following components:

- context
- anticipated outcomes/preconditions modelled in a causal pathway
- process/sequence of interventions (activities) required to achieve change(s)
- assumptions about how these changes may happen and
- diagram and narrative summary (12).



FURTHER  
READING

### Conceptual frameworks

Earp JA, Ennett ST. Conceptual models for health education research and practice. *Health Educ Res.* 1991;6(2):163–71. doi:10.1093/her/6.2.163. (16)

### Results frameworks

Performance monitoring and evaluation TIPS: building a results framework. No 13, second edition. Washington (DC): United States Agency for International Development (USAID); 2010 (<https://www.ndi.org/files/Performance%20Monitoring%20and%20Evaluation%20Tips%20Building%20a%20Results%20Framework.pdf>). (10)

### Theory of change

Vogel I. Review of the use of “Theory of Change” in international development: review report. London: United Kingdom Department of International Development; 2012 ([http://r4d.dfid.gov.uk/pdf/outputs/mis\\_spc/DFID\\_ToC\\_Review\\_VogelIV7.pdf](http://r4d.dfid.gov.uk/pdf/outputs/mis_spc/DFID_ToC_Review_VogelIV7.pdf)). (12)



## Part 2c: Setting the stage: selecting indicators for digital health interventions



### HOW WILL THIS SECTION HELP ME?

This section will:

- ✓ Demonstrate how to select appropriate indicators to adequately monitor and evaluate digital health interventions
- ✓ List illustrative indicators determined to be useful for monitoring and evaluation (M&E) of digital health interventions
- ✓ Provide key categories of indicators to be considered for conducting M&E of digital health interventions.



### KEY TERMS

**Functionality (also referred to as functional suitability):** A “characteristic that represents the degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions” (21). In this Guide, functionality refers to the ability of the digital health system to support the desired digital health intervention.

**Indicator:** “A quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor” (14).

**Usability:** The “degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (22).

**Users:** The individuals who directly utilize the technology using their digital devices, either to deliver health services (e.g. community health workers, district managers, clinicians) or to receive services (i.e. clients, patients).

Development of a set of indicators to measure how well programme activities have been implemented and their impact on health outcomes is central to programme monitoring and evaluation (M&E). This chapter discusses various considerations for the selection of indicators, and presents a generic listing of indicators judged to be useful for M&E of digital health interventions.

WHO defines an **indicator** as “a quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor” (14). Each intervention activity (also referred to as programme activities in standard evaluation frameworks) should have at least one measurable indicator, with no more than 10–15 indicators for each programmatic area. Indicators can be qualitative (e.g. availability of a clear, organizational mission statement) or quantitative (i.e. expressed as numbers or percentages).

The SMART criteria below provide some guidance for constructing indicators.

- **S = Specific:** The indicator must be specific about what is being measured, from whom the data will be collected and when.
- **M = Measurable:** The indicator must be quantifiable. Avoid the use of subjective terms such as “good quality” or “accessible” in defining the indicator since these may be interpreted differently across regions, professions and individuals.

- **A = Attainable:** The indicator should be attainable with the available budget, time and human resources.
- **R = Relevant:** The indicator should be relevant to the context, and specific to the needs of the programme or intervention being evaluated.
- **T = Time-bound:** The indicator should be time-specific, based on the time frame of the health programme.

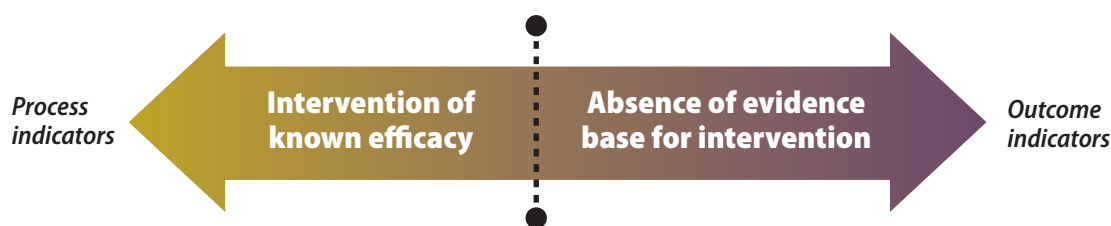
## Approach for selection of indicators for evaluating digital health interventions

The selection of specific indicators for programme assessment depends largely on the goals and objectives of that programme, but there are certain general guiding principles. First, indicator selection should be based on close alignment with digital health intervention aims and priorities, and with practical considerations in terms of the context and availability of resources. The indicators should also align with the claims (considerations for identification of claims are described in Chapter 2, Part 2a). For each project, the choice of indicators must be linked to what the projects aims to do, who the consumers of the data are (i.e. stakeholders such as donor agencies and the government), and what kinds of decisions need to be made based on the data (e.g. validating the digital health strategy, improving the implementation process).

Typically in global health programmes, evidence of impact or direct improvements in health outcomes is the benchmark of the intervention's validity. However, with digital health interventions, the focus thus far has been on the use of digital health systems to improve upon and streamline the delivery of existing health services, to bring about improved population coverage and quality of services provided. Claims have also been made that the use of digital health interventions supports positive health behaviours and reduces the costs of service delivery by creating effective channels for data transfer and communication. Most often, a digital health intervention serves as an adjunct to or a catalyst for an existing intervention that is known to be effective. In such cases, where it is not feasible to measure impact, proxy indicators and process indicators can be used to assess the effectiveness of digital health interventions.

Figure 2.9 illustrates a moving “barometer” of the kinds of indicators necessary for evaluating a digital health intervention. If a digital health intervention project is using a novel approach, where there is no strong evidence base or precedent supporting the underlying intervention, then the barometer for the indicators would move to the right, relying more heavily on outcome indicators. However, when the evidence base is already robust for the underlying intervention (e.g. vaccines, medication regimens, clinical care), then the barometer moves to the left, focusing on process/functionality indicators. Outcomes, in the latter case, have already been established through prior research involving the same underlying intervention and the new challenge for the digital health intervention would be to improve reach/coverage or, possibly, timeliness of the delivery of that intervention.

**Figure 2.9. “Barometer” for selection of digital health indicators**

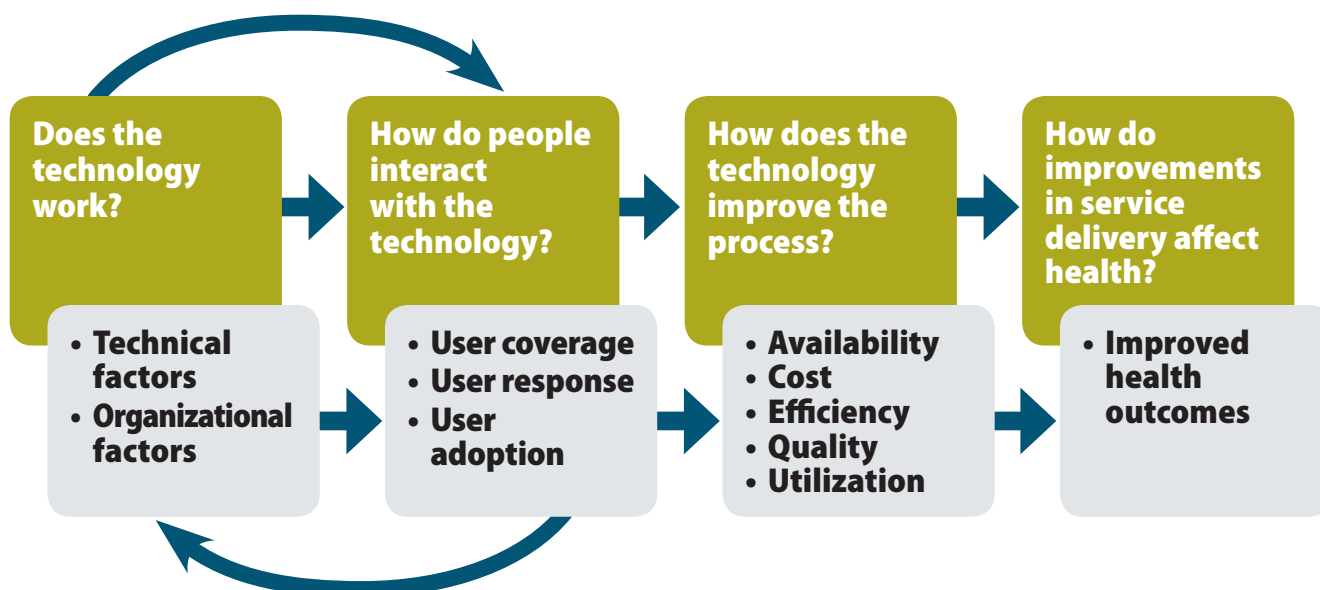


## Overview of classification of digital health indicators

The framework presented in Figure 2.10 identifies key areas through which digital health interventions achieve results. Indicators should be aligned with the overarching research question(s), which are presented in the coloured boxes along the top. The framework in the figure provides a basis for the assessment of digital health intervention performance in a number of areas, including: (a) the technical and organizational aspects of the digital health system; (b) the target audience's usage of and response to the digital health intervention; (c) the intervention's success in addressing constraint areas for the process of health service delivery; and (d) the effect on improving health outcomes.



**Figure 2.10. Categorization of digital health indicators**



- The first question of the framework – Does the technology work? – relates to assessment of the inputs for developing a digital health system (i.e. the technology and application), in addition to an assessment of the feasibility of the digital health intervention.
- The second question – How do people interact with the technology? – covers service output measures intended to capture and assess the immediate results of the intervention (23). Additionally, it captures usability measures that will help to quantify how the users interact with the system.
- The third question – How does the technology improve the process? – captures the effect of the digital intervention on service utilization outputs or the extent to which clients use the service, and intermediate population-level outcomes. It also captures process and early outcome indicators.
- The fourth question – How do improvements in service delivery affect health? – captures long-term outcomes and impact.

In the following sections, we describe each of these components in greater detail, and identify 10–15 sample indicators in each category. The indicators are generic and are not intended to be exhaustive. Priority indicators can be selected based on the relevance to the digital health intervention and modified to reflect the specific objectives of the intervention.

## Functionality – Does the technology work?

The indicators in this group seek to determine:

- Technology design – Does the technology perform its intended functions effectively?
- Technology adaptation to the local context – Is the technology effectively adapted to the local context in terms of language, literacy, modifications for network coverage, etc.?

### Box 2.6. The PRISM framework

The Performance of Routine Information System Management (PRISM) framework identifies key areas that affect health information systems (HIS) and provides structured methods for assessing HIS performance (24). In addition to technical factors, PRISM also focuses on organizational factors (i.e. the health services delivery system, which may include factors such as inadequacies in financial and human resources, management support, supervision and leadership) and behavioural factors (e.g. user demand, motivation, confidence and competence in using the system), recognizing that even the most sophisticated technology has limitations and will fail to achieve intended results without the necessary enabling context and factors (25).

## Technical factors

Technical factors for assessing digital health systems include factors that relate to specialized knowledge and application of software development, information technology (IT) for data processing, data security protocols and the relevance and management of the system in the context of the intervention programme (26). In order to assess whether the digital health system is appropriate to the context, it is most important to assess and record infrastructure availability, such as mobile network coverage. The section on technical factors in Table 2.4 lists sample domains and indicators for measuring these factors, covering issues ranging from access to skilled local staff for technical support and maintenance, to local levels of literacy and ability to use the relevant mobile phone functions.

**Table 2.4. Sample domains and indicators: Does the technology work?**

Metric area	Indicators
<b>TECHNICAL FACTORS</b>	
<b>Connectivity</b>	% of target population with mobile phone signal at time of interview
<b>Power</b>	% of target population with current access to a power source for recharging a mobile device
<b>Skilled local staff</b>	% of digital health interventions with access to local technical support for troubleshooting
	% of users with access to local technical support systems for troubleshooting
<b>Maintenance</b>	% devices that are not currently operational (misplaced/broken/not working)
<b>Functionality</b>	% of mobile devices that are operational in the language of the users
	% target population who are literate in the language used by the digital health intervention
	% of target population who report ever using short message service (SMS) capabilities
	% of data fields from original paper-based system that are captured by the technology
<b>ORGANIZATIONAL FACTORS</b>	
<b>Training</b>	No. hours of initial training on the use/deployment of the technology attended by programme staff
	No. hours of refresher training on the use/deployment of the technology attended by programme staff

## Qualitative approaches to assess technical factors

In addition to the above-mentioned criteria, other considerations factor into the development and continuous improvement of a digital health system. Documentation of certain qualitative measures would promote programmatic and contextual understanding, especially during the pre-prototype and prototype stages of development.

For example:

- **Needs assessment:** Does the system address an identified public health need?
- **Software considerations:** Does the software comply with current industry standards?

The Software product Quality Requirements and Evaluation (SQuaRE) criteria further inform the process of identification of technical factors for evaluation (27). Software quality considerations are key to ensuring that the digital health system meets the industry standards, adequately addresses users' needs and accounts for different local area context/ environments. According to ISO/IEC 25000:2014, software quality should be assessed while the product is under development (internal software quality), during testing in a simulated environment (external software quality) and when the product is in use (28).

The development of the digital health system is an iterative process, repeatedly reviewing and making adjustments based on changes in the stability of the software and hardware, its application and use in different environments, and user feedback, in order to further improve the system. However, often the system or technologies used are based on repurposing of existing technologies – in which case end-user inputs on the existing technology should be incorporated at the earliest possible stage. Refer to Chapter 3 on monitoring for further details on assessing technical functionality and stability.

## Organizational factors

These factors relate to the organizational context within which the digital health system is being used as part of an intervention – regardless of whether it is hosted by a private or public entity. When assessing a digital health intervention, indicators should cover organizational factors such as inadequacies in training, supervision and/or leadership as relevant to the adoption of the digital health intervention by intended users, as well as the financial resources the organization has invested in the development and maintenance of the system. Table 2.4 lists sample indicators aimed at measuring the number of hours programme staff at the organization have spent in training on the use and deployment of the digital health system.

## Data sources

While specific data sources will vary by project, data may be drawn from primary and secondary sources. Primary sources of data are likely to include system-generated data and data collected through quantitative and/or qualitative surveys, including surveys of users of the technology. Existing regional or national telecommunications reports might provide good data as a basis for assessing the existing connectivity and infrastructure. Organizational indicators should be captured on a routine basis as part of internal activity reporting (e.g. data on the number of training activities held).

## Usability – How do people interact with the technology?

The success of a digital health intervention, including the level of adoption by users in the target population, is dependent on the end-users' interaction with the technology and their belief/opinion that use of the technology will benefit their health or finances (or those of their clients, in the case of health workers). This group of indicators addresses the assessment of the response of the end-users to the digital health intervention.

Output indicators can be used for multiple functional areas essential to support programme activities. These areas include, but are not limited to, programme management (e.g. number of managers trained), advocacy (e.g. number of advocacy meetings held), behaviour change communication, availability of commodities, and policy. Indicators for functional outputs would capture the number/quantity of activities conducted in each area of service delivery (e.g. the number of behaviour change communication messages). Indicators for service outputs measure the quantity/quality of services, including the content of the care or information provided to the target population of users (e.g. quality of care, client satisfaction). Table 2.5 identifies key indicators in this category.

Behavioural factors may influence demand for and use of the digital health intervention, including confidence, satisfaction and competence in using the system. One of these factors is the end-users' ability to use the system; therefore, it may be of interest to assess whether the technology platform (the digital health system) has taken this ability into account. This ability is reflected in the rates of use of the digital health system, including frequency of data upload/submission and quality of data entry. The "user" refers to the individuals who directly utilize the technology on their digital devices, either to deliver health services (e.g. community health workers, district managers, clinicians) or to receive services (i.e. clients, patients).

Indicators in this category delve into the following questions:

- User coverage:
  - ✓ Has the digital health system been widely adopted? This may be measured as the percentage of the target population who have adopted (i.e. become users of) the technology.
- User response:
  - ✓ Do the users find the technology easy to use?
  - ✓ Do the users find the health information received through the digital health intervention useful?
- User adoption:
  - ✓ Are the users able to communicate with the digital health system as intended? Are they responsive to the information received through the system?

Note that adoption or coverage rates may have also been affected by input-level factors discussed previously in the earlier section on functionality – Does the technology work?

**Table 2.5. Usability indicators: How do people interact with the technology?**

Metric area	Indicators
<b>User coverage</b>	% of users who demonstrate proficiency in use of the digital health system % of intended users observed using the digital health system over reference period No. transmissions sent by intended users over reference period
<b>User response</b>	% of users who rate the digital health system as “easy to use” % of users who rate the digital health system as “transmits information as intended” % of users who report satisfaction with the content of health information received via the digital health system % of users motivated/intending to use the digital health system
<b>User adoption</b>	% of messages transmitted via the digital health system that are responded to appropriately <sup>a</sup> by end-user over reference period No. messages/forms/amount of data transmitted by end-user via the digital health system within reference period % of data fields/forms that are left incomplete over reference period

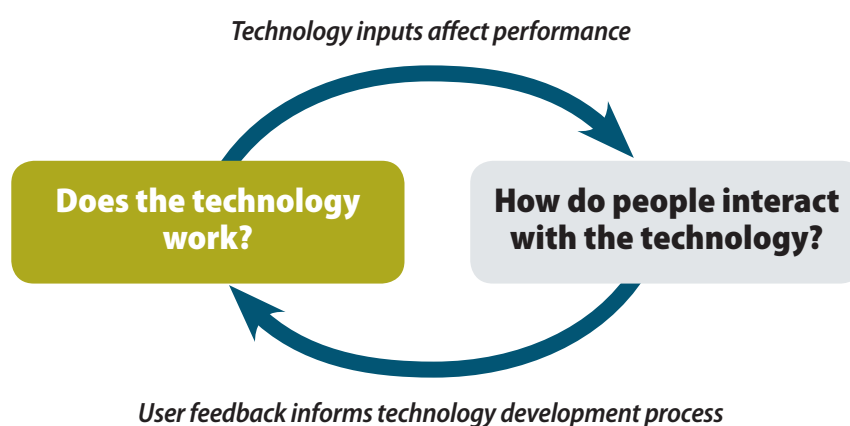
a “Appropriately” could refer to completion of intended action to reflect that the message has been read, e.g. acknowledgement of message.

## Data sources

Digital health systems offer unique opportunities to capture several of these output indicators, especially the functional output indicators, by using routinely collected back-end data. Not only can routine monitoring data, which was traditionally captured on paper records, now be instantly digitized, but it can also be automatically analysed and presented on interactive dashboards to make data use easy. Data on indicators such as client satisfaction can be captured through user surveys.

## Feedback loop

Implicit in components (or questions) 1 and 2 of indicator categorization is a feedback loop, as shown in Figures 2.10 and 2.11. From a technical perspective, any performance feedback derived from the information about end-user adoption and satisfaction rates would loop around to further inform the software development process and determine the quality of the technological inputs. This, in turn, would affect the performance of the revised version of the digital health system among end-users, making the technology development an iterative process. In the field of engineering, this process is referred to as a “spiral model of software development and enhancement”. It entails a process of iterative prototyping and product enhancement followed by the end-user reviewing the progress (29).

**Figure 2.11. Feedback loop: an iterative process of development**

## Process improvement – How does the technology improve service delivery?

The potential benefits of digital health interventions include improved efficiency of health service delivery and data collection, and the ability to provide and exchange information on demand, facilitating communication across different levels of the health system and among providers (30). This group of indicators makes the leap from coverage rates of the digital health intervention itself, to the measurement of service utilization outputs and early- to intermediate-stage health outcomes across the three levels of health service delivery: client, provider and health system. As depicted in Figure 2.12, the development of indicators at this stage is based on the identification of key digital health intervention areas as they address the “constraints” at each level of service delivery.

Indicators at this level must be focused on the need to evaluate the effectiveness of a digital health intervention in addressing the constraints of coverage and scale (availability), costs of delivery, technical efficiency, quality and utilization of health services.

**Figure 2.12. How does the technology improve the process? Addressing “constraints” across levels of service delivery**



### Measurement at each level of the health system

A digital health intervention may operate at one or more levels of the health system. For example, programmes targeted at behaviour change communication, such as the Mobile Alliance for Maternal Action (MAMA) or m4RH, are largely focused at the client level and so indicators may only be needed at this level. Programmes such as cStock, on the other hand, work with providers and clients at several levels of the health system to reduce stock-outs of drugs, and would need to measure indicators at each of the three levels: client, provider and health system.

#### • Client-level indicators

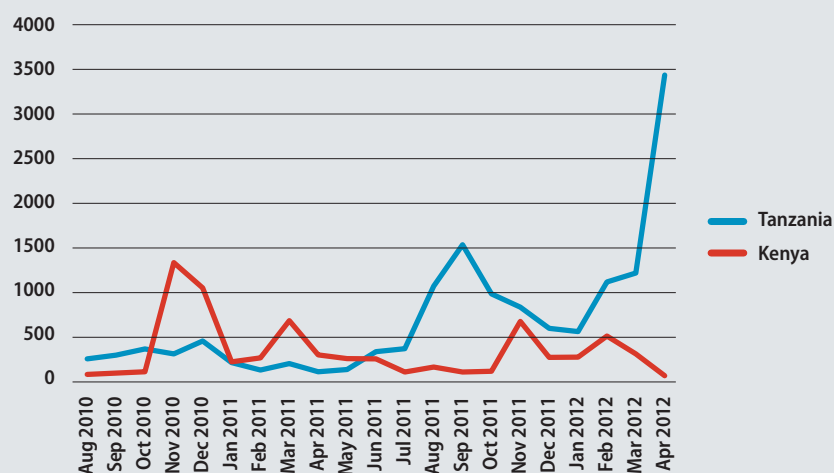
“Client” refers to the person who benefits from the digital health intervention in a way that directly affects their health. This may also include the family members of the direct recipient of the health services. Client-level measures seek to assess the direct outcomes of the digital health intervention as experienced by these beneficiaries. Table 2.6 presents critical indicators at this level.

## m4RH case study: Monitoring intervention output to understand coverage and marketing approach

The Mobile for Reproductive Health (m4RH) project, developed by FHI 360, comprises a set of evidence-based text messages, which users in Kenya and the United Republic of Tanzania can access via their mobile phones, providing information about family planning methods. Users interested in accessing m4RH messages can text a short code to m4RH to start receiving the informational messages. Each time the user “pings” the digital health system, it is registered on the back-end. These back-end data can be used to monitor the coverage and usability of the digital health intervention.

Figure 2.13 shows monitoring data captured on the m4RH back-end for the indicator “number of unique clients who contacted the m4RH system per month”. It allows the programme implementers to answer programmatic questions such as What percentage of our target population are we reaching? and How do promotional activities affect coverage of m4RH?

**Figure 2.13. m4RH monitoring data: number of unique clients who contacted the system, per month, August 2010 to April 2012**



Source: FHI 360, undated (31).

### • Provider-level indicators

Provider refers to any provider of health services, including front-line health workers, clinic staff and managers. The sample provider-level indicators presented in Table 2.7 are disaggregated as proportions of all providers and averages per provider, over a reference period.

### • Health-system-level indicators

The WHO Framework for strengthening of health systems identifies six building blocks: service delivery; health workforce; health information systems; access to essential medicines; financing; and leadership/governance (32). Table 2.8 identifies generic indicators that can be applied to assess the effect of the digital health intervention on each of these building blocks.

## Constraint considerations for the recommended indicators

Four “constraint” categories – cost, technical efficiency, quality and utilization – are discussed separately below to help the reader think about indicators relating to the different constraints that their intervention might address. However, it should be understood that these categories are not always mutually exclusive. For example, depending on the programme objectives, quality of care may include availability of information, affordability, access and technical efficiency.

### • Costs

At the client level, costs refer to the direct costs (e.g. fees paid for health services) and indirect costs (e.g. transportation costs, opportunity costs and loss of income [33]) incurred by the client. When observed and captured over time in the target population, it is envisioned that digital health interventions might lead to cost savings due to care seeking behaviour that is increasingly more timely and appropriate. Therefore, for the purpose of developing an evidence base for digital health interventions, it is critical not only to measure the costs incurred directly through implementation of the digital health intervention, but also the costs averted at each level as a result of clients receiving health services remotely or more timely identification of an illness, thus avoiding costs associated with the progression of the disease. Cost indicators for digital health interventions should be disaggregated by each level to include specific areas where changes in costs are expected.



**Table 2.6. Client-level indicators: How does the technology improve the implementation process?**

Digital health metric	Indicators
<b>Efficiency</b>	<p>No. minutes (reported or observed) between digital health system prompt received about intervention X and seeking care from provider</p> <p>No. in-person consultations with qualified health-care providers about intervention X by target clients as a result of accessing required services using digital health intervention over reference period<sup>a</sup></p> <p>No. days duration of illness episode</p>
<b>Quality</b>	<p>No. minutes spent with a health-care provider in relation to health intervention X at the last visit</p> <p>% of messages received through digital health intervention that clients are able to recall about intervention X during client exit interviews</p> <p>% of target clients who report correctly adhering to prescribed care protocol in relation to intervention X</p>
<b>Utilization</b>	<p>% of emergency events where the digital health system was used by patients to expedite treatment over reference period</p> <p>% of target clients who report receiving health information about intervention X via their mobile phone within reference period</p> <p>% of target clients who report contact<sup>b</sup> with a qualified health-care provider using the digital health system in relation to intervention X over reference period</p> <p>% of target clients who report adequate<sup>c</sup> knowledge about signs and symptoms for which they should seek care in relation to intervention X</p> <p>% of target clients who report adequate<sup>c</sup> knowledge about the health issue relevant to intervention X</p>
<b>Costs</b>	<p>% changes in reported client out-of-pocket payments for illness management over reference period (through managing the illness by phone-based consultation instead of visiting a health-care facility, e.g. travel cost)<sup>d</sup></p>

X: Insert name of the specific health intervention targeted by the digital health system.

<sup>a</sup> Requires collection at multiple time points to yield estimates of “averted” incidences.

<sup>b</sup> “Contact”: To be determined based on digital health intervention medium of health service delivery. Could include telephonic consultation, home visit by health worker, or clinic visit by patient where the use of the digital health intervention has played a role in the receipt of services.

<sup>c</sup> “Adequate” could be defined by programme intervention, e.g. % of target clients who know three pregnancy danger signs.

<sup>d</sup> Composite indicator – could be sub-categorized into individual components of interest where cost savings are intended.

**Table 2.7. Provider-level indicators: How does the technology improve the implementation process?**

Digital health metric	Indicator
<b>Efficiency</b>	<p>No. minutes (reported or observed) for last client counselling about intervention X using digital health system</p> <p>No. minutes or hours (reported or observed) spent on health record-keeping about intervention X over reference period</p> <p>No. minutes (reported or observed) used per individual health worker over reference period to transmit data relating to intervention X from community-based logs to health-care facility-based information systems</p> <p>No. minutes (reported or observed) taken per individual health-care provider over reference number of events between identification of an adverse event and provision of care (intervention X), across levels of a health system</p> <p>No. minutes (reported or observed) used per individual health worker to report important adverse events (e.g. stock-outs)</p>
<b>Quality</b>	<p>% of health workers who report adequate<sup>a</sup> knowledge of the health issue relevant to intervention X</p> <p>% of care standards relating to intervention X observed to be met using the digital health intervention during a client–provider consultation</p> <p>% of providers observed to be using the digital health intervention during their patient consultations</p>

Digital health metric	Indicator
<b>Utilization</b>	% of targeted health workers who use the digital health system in relation to intervention X through their mobile phones over reference period
	% of health workers observed to use the digital health system during their last client contact
	% of health workers who use the digital health system to connect <sup>b</sup> with medical staff to receive real-time clinical information and decision support
	No. clients (average or total) attended <sup>c</sup> by a health worker using the digital health system over reference period
<b>Costs</b>	Amount of cost savings (estimated) due to improvement in service delivery/efficiency <sup>d</sup> /other factors.

X: Insert name of the specific health intervention targeted by the digital health system.

<sup>a</sup> "Adequate" could be defined by programme intervention, e.g. % of target health workers who know three pregnancy danger signs.

<sup>b</sup> "Connect" could be via phone call, e.g. community health workers might call health supervisors for suspected complication and received decision support via phone call or other digital health supported means from a high-level provider.

<sup>c</sup> "Attended" could be via phone call or personal home visit or other modes of communication using digital health intervention.

<sup>d</sup> Composite indicator derived through monetizing time savings for administrative functions.

**Table 2.8. Health-system-level indicators: How does the technology improve the implementation process?**

Digital health metric	Indicators
<b>Efficiency</b>	No. minutes (cumulative) over reference period for all health workers in a health-care facility using digital health system to enter data related to intervention X <sup>a</sup>
	No. minutes (cumulative) over reference period for all health workers to transmit data about intervention X from community-based logs to health-care facility information systems
	No. minutes (cumulative) over reference number of events between identification of an adverse event and provision of care (intervention X), across levels of a health system
	No. days over reference period for which a health-care facility reports stock-out of a commodity essential for provision of intervention X
<b>Quality</b>	No. health workers observed to be providing clinical services related to the digital health intervention
	% change in reported stock-out events of a commodity essential for provision of intervention X over reference period <sup>b</sup>
	% change in data entry errors over reference period <sup>b</sup>
	% of target health workers who receive initial training on using the digital health system to deliver intervention X
	% of target health workers who receive refresher training on using the digital health system to deliver intervention X (initial and refresher training)
<b>Utilization</b>	No. clients seeking intervention X over reference period
	% of clients in a specified area who receive intervention X through the digital health system over reference period
	% of target population who have access to intervention X over reference period
	% of health-care facilities in a target geographical area that use the digital health intervention
	No. clients seeking intervention X at health-care facility using the digital health system
<b>Costs</b>	% change in costs of transporting paper forms and manual data entry over reference period <sup>b</sup>
	% change in costs of human resources for data entry <sup>b</sup>
	% change in costs associated with timely and appropriate management of illness <sup>b</sup>
	% changes in reported client out-of-pocket payments for management of illness <sup>b</sup>
	Total population-level savings in out-of-pocket payments attributed to timely and appropriate care seeking <sup>b</sup>

X: Insert name of the specific health intervention targeted by the digital health system.

<sup>a</sup> Aggregated facility-level indicator (corresponding indicator at provider level is disaggregated).

<sup>b</sup> Assumes data collection at two points – before and after the implementation of the digital health intervention.

At the health system level, it is of interest to measure not only the achievement of superior clinical outcomes, but also the achievement of these outcomes at a reduced cost. Assessing the costs and savings relating to staff and client time would entail assigning a monetary value to units of time in order to monetize the anticipated time gains that may result from employing a digital health intervention. Additional considerations for costs can be derived from an understanding of the areas of increased operational efficiency, e.g. costs averted as a result of timely identification of an emergency, human resources costs averted due to reduced need for manual data entry.

Cost-related data can be collected from programme records and special surveys. For additional details on appropriate methods for collecting and managing cost data, refer to Chapter 4, Part 4b.

- **Technical efficiency**

An intervention is said to be technically efficient if a particular output can be obtained with less input; in other words, use of the available resources is maximized (34). At the client level, recommended indicators that measure technical efficiency include those measuring savings in the time it takes for the patient to receive care, reduced duration of illness and reduced need to consult a facility-based health-care provider.

At the provider level, technical efficiency refers to effects such as changes in a provider's allocation of time to clinical versus administrative functions, and changes in the time taken to respond to an adverse event. Monetization of such time-based technical efficiency indicators would yield a measure of cost savings. At the health system level, efficiency indicators show the cumulative time savings for all the health-care providers who are part of that system.

Collection of data on technical efficiency typically requires additional surveys. Where a digital health intervention involves delivery of a service by a provider using a mobile device, the back-end data may have timestamps, which can be used for measures of technical efficiency.

- **Quality**

Quality of care can be measured through the three dimensions of structure, process and outcome (35). Structure includes attributes of material resources (such as facilities, equipment and finances), as well as organizational structure. Process includes the activities that are carried out in the course of providing care. Outcome refers to the effect of the services on health. The definition of "quality" may also include dimensions of effectiveness, efficiency, acceptability and equity. Improvements in service quality at the client level may result from improved efficiency and knowledge at the service provider level, as well as self-reported response to health reminders received through a digital health intervention. For quality indicators at a provider level, evidence of knowledge and improved ability to provide services serve as proxy indicators. Changes in operational efficiency and cumulative quality gains yield quality measurements at the health system level. Depending on how quality is defined by the project, data might be collected during routine procedures (e.g. comprehensive counselling can be assessed using back-end data collected as part of routine service delivery using a mobile job aid), or may require additional surveys.

- **Utilization**

Utilization is a function of availability of services, user needs, perceptions and beliefs. The key question to be answered is: Did the digital health intervention improve utilization of health services by clients? At the client level, this refers to the availability (coverage) and accessibility of health services. Coverage is a complex concept to measure, as it is influenced by both demand-side and supply-side factors. Health services in this context could include either in-person service delivery or provision of health-related information. At the provider level, coverage refers to the availability of and access to training and decision-support services for community- and facility-based providers. At the health system level, indicators of utilization capture aggregated coverage of services based on hospital and community records.

A distinction is made here between individual-level data, which are collected directly from the client at the community level, and facility-level data, collected from health-care facility records. The health-care facility-level indicators are listed in Table 2.8 and can serve to triangulate the information collected from direct interviews with end-users, as utilization data collected from end-users may be subject to recall bias. Utilization data may be abstracted from health-care facility records and back-end system data, or collected purposively using additional surveys.

## Health outcomes – How do improvements in service delivery affect health outcomes?

The fourth question of the framework addresses the health outcomes as a result of improvements in service delivery. The distinction in digital health evaluation from traditional evaluation is that there is not always a need to evaluate health outcomes as direct effects of the digital health intervention. As depicted in Figure 2.10, the rationale for the use of outcome indicators in the evaluation of a digital health intervention is the absence of prior research validating the health-related intervention. For interventions with known efficacy based on prior research, the focus of the digital health intervention evaluation can be limited to the evaluation of the process and early outcomes, based on the types of indicators presented in Tables 2.4–2.8. For example, an evaluation of a digital health intervention for vaccination reminders should focus on measuring the percentage of children who received timely vaccination as a result of the digital prompt (e.g. text message) and need not seek to measure the direct impact of vaccination on the rate of child mortality from the disease they are vaccinated against, since the effectiveness of timely vaccination will have been established through prior research.



### FURTHER READING

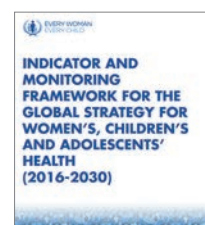
To avoid duplication of work, this chapter has focused on input, process and early outcome indicators specific to digital health strategies. While intermediate outcome and impact indicators are important, they have not been listed in this chapter since there are a number of other existing indicator databases that provide these in detail. Suggestions of repositories for relevant indicators include Lives Saved Tools (LiST), MEASURE Evaluation, Countdown to 2015, UNFPA Toolkits, USAID Maternal and Newborn Standards and Indicators Compendium, among others. Specific standardized indicator databases for outcome and impact measurement are presented in the box of further resources, below.

### Resources for standardized list of indicators

#### 1. Reproductive, maternal, newborn, child and adolescent health (RMNCAH) indicators

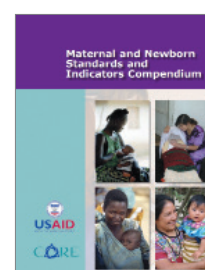
Every Woman Every Child – Indicator and monitoring framework for the Global Strategy for Women's, Children's and Adolescents' Health (2016–2030)

[http://www.everywomaneverychild.org/images/content/files/EWEC\\_INDICATOR\\_MONITORING\\_FRAMEWORK\\_2016.pdf](http://www.everywomaneverychild.org/images/content/files/EWEC_INDICATOR_MONITORING_FRAMEWORK_2016.pdf)



Maternal and newborn standards and indicators compendium

[http://www.coregroup.org/storage/documents/Workingpapers/safe\\_motherhood\\_checklists-1.pdf](http://www.coregroup.org/storage/documents/Workingpapers/safe_motherhood_checklists-1.pdf)



Demographic and Health Surveys (DHS) survey indicators – Maternal and child health

<http://www.dhsprogram.com/data/DHS-Survey-Indicators-Maternal-and-Child-Health.cfm>





## 2. HIV programmes

National AIDS programmes – A guide to indicators for monitoring and evaluating national HIV/AIDS prevention programmes for young people

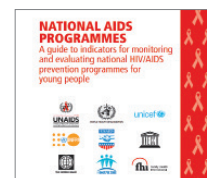
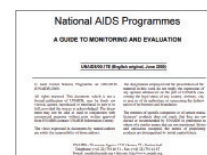
[http://www.unaids.org/sites/default/files/media\\_asset/jc949-nap-youngpeople\\_en\\_1.pdf](http://www.unaids.org/sites/default/files/media_asset/jc949-nap-youngpeople_en_1.pdf)

HIV/AIDS Survey Indicators Database – Behavioural and outcome indicators

[http://hivdata.measuredhs.com/ind\\_tbl.cfm](http://hivdata.measuredhs.com/ind_tbl.cfm)

National AIDS programmes – A guide to monitoring and evaluation

<http://hivdata.measuredhs.com/guides/unaidsguide.pdf>



## 3. Malaria programmes

Demographic and Health Surveys (DHS) Malaria Indicator Survey (MIS)

<http://dhsprogram.com/What-We-Do/Survey-Types/MIS.cfm>

The President's Malaria Initiative

[http://www.pmi.gov/docs/default-source/default-document-library/tools-curricula/pmi\\_indicators.pdf?sfvrsn=4](http://www.pmi.gov/docs/default-source/default-document-library/tools-curricula/pmi_indicators.pdf?sfvrsn=4)



## 4. Health service delivery indicators

WHO Health service delivery indicators, including service provision assessment (SPA) and quality of care

[http://www.who.int/healthinfo/systems/WHO\\_MBHSS\\_2010\\_section1\\_web.pdf](http://www.who.int/healthinfo/systems/WHO_MBHSS_2010_section1_web.pdf)



MEASURE Evaluation Service delivery – Quality of care/service provision assessment

[http://www.cpc.unc.edu/measure/prh/rh\\_indicators/crosscutting/service-delivery-ii.h.2](http://www.cpc.unc.edu/measure/prh/rh_indicators/crosscutting/service-delivery-ii.h.2)



## 5. MEASURE Evaluation Summary list of indicators

Cross-cutting indicators, including women's and girls' status empowerment, health systems strengthening, training, commodity and logistics, private sector involvement, behaviour change communication, access, quality of care, gender equity, and programmatic areas (maternal, newborn and child health, family planning, safe motherhood and post-abortion care, HIV/AIDS/STIs, adolescent health, gender-based violence, and male involvement in reproductive health)

[http://www.cpc.unc.edu/measure/prh/rh\\_indicators/indicator-summary](http://www.cpc.unc.edu/measure/prh/rh_indicators/indicator-summary)



## References

1. Crafting your value proposition. In: MaRS [website]. Toronto: MaRS Discovery District; 2012 (<https://www.marsdd.com/mars-library/crafting-your-value-proposition/>, accessed 21 March 2016).
2. Business Model Design. In: MaRS [website]. Toronto: MaRS Discovery District; 2012 (<https://www.marsdd.com/mars-library/business-model-design/>, accessed 21 March 2016).
3. CommCare [website]. Dimagi; 2016 (<https://www.commcarehq.org/home/>, accessed 21 March 2016).
4. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol*. 1999;28(1):10–8.
5. Mehl G, Labrique A. Prioritizing integrated mHealth strategies for universal health coverage. *Science*. 2014;345(6202):1284–7. doi:10.1126/science.1258926.
6. Bryson JM, Patton MQ, Bowman RA. Working with evaluation stakeholders: a rationale, step-wise approach and toolkit. *Eval Program Plann*. 2011;34(1):1–12. doi:10.1016/j.evalprogplan.2010.07.001.
7. Bryson JM. What to do when stakeholders matter? *Public Manag Rev*. 2004;6(1):21–53. doi:10.1080/14719030410001675722.
8. Gable C, Shireman B. Stakeholder engagement: a three-phase methodology. *Environ Qual Management*. 2005;14(3):9–24. doi:10.1002/tqem.20044.
9. Monitoring and evaluation frameworks (3 parts). In: Virtual Knowledge Centre to End Violence against Women and Girls [website]. UN Women; 2012 (<http://www.endvawnow.org/en/articles/335-monitoring-and-evaluation-frameworks-3-parts.html>, accessed 3 May 2016).
10. Performance monitoring and evaluation TIPS: building a results framework, No. 13, second edition. Washington (DC): United States Agency for International Development (USAID); 2010 (<https://www.ndi.org/files/Performance%20Monitoring%20and%20Evaluation%20Tips%20Building%20a%20Results%20Framework.pdf>, accessed 9 May 2016).
11. Monitoring and evaluation: some tools, methods, and approaches. Washington (DC): The International Bank for Reconstruction and Development/The World Bank; 2004.
12. Vogel I. Review of the use of “Theory of Change” in international development: review report. London: United Kingdom Department of International Development; 2012 ([http://r4d.dfid.gov.uk/pdf/outputs/mis\\_spc/DFID\\_ToC\\_Review\\_VogelV7.pdf](http://r4d.dfid.gov.uk/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf), accessed 9 May 2016).
13. Glossary of selected planning, monitoring and evaluation terms. Measurement, Learning and Evaluation (MLE) Project, Urban Reproductive Health Initiative; 2013 (<https://www.urbanreproductivehealth.org/toolkits/measuring-success/glossary-selected-planning-monitoring-and-evaluation-terms>, accessed 4 May 2016).
14. WHO evaluation practice handbook. Geneva: World Health Organization; 2013 ([http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf), accessed 4 May 2016).
15. Measuring success toolkit: frameworks. In: Toolkits by K4Health [website]. K4Health; 2016 (<https://www.k4health.org/toolkits/measuring-success/frameworks>, accessed 22 April 2016).
16. Earp JA, Ennett ST. Conceptual models for health education research and practice. *Health Educ Res*. 1991;6(2):163–71. doi:10.1093/her/6.2.163.
17. Using mobile technology to strengthen maternal, newborn, and child health: a case study of MOTECH’s five years in rural Ghana. Washington (DC): Grameen Foundation; 2015 (<https://grameenfoundation.app.box.com/v/motechghanareport>, accessed 5 May 2016).



18. mHealth for MNCH impact model. In: GSMA [website]. GSMA; 2014 (<http://www.gsma.com/mobilefordevelopment/programme/mhealth/mhealth-for-mnch-impact-model>, accessed 4 May 2016).
19. MomConnect: the NDOH initiative. Presented at the ECD Knowledge Building Seminar. 25–26 November 2016. South Africa, National Department of Health (NDOH); 2015 ([http://www.unicef.org/southafrica/SAF\\_kbs2015\\_4\\_1MomConnectHowitworkswhatitachieves.pdf](http://www.unicef.org/southafrica/SAF_kbs2015_4_1MomConnectHowitworkswhatitachieves.pdf), accessed 22 April 2016).
20. Malawi. In: Supply Chains 4 Community Case Management (SC4CCM) [website]. Arlington (VA): JSI Research & Training Institute, Inc.; 2016 (<http://sc4ccm.jsi.com/countries/malawi/>, accessed 21 April 2016).
21. Functional suitability. In: ISO 25000 software product quality [website]. ISO 25000; 2015 (<http://iso25000.com/index.php/en/iso-25000-standards/iso-25010/58-functional-suitability>, accessed 9 May 2016).
22. Usability. In: ISO 25000 software product quality [website]. ISO 25000; 2015 (<http://iso25000.com/index.php/en/iso-25000-standards/iso-25010?limit=3&start=3>, accessed 9 May 2016).
23. Types of data and indicators. In: Family Planning and Reproductive Health Indicators Database, MEASURE Evaluation Population and Reproductive Health Project [website]. Chapel Hill (NC): MEASURE Evaluation; 2016 ([http://www.cpc.unc.edu/measure/prh/rh\\_indicators/overview/types-of-indicators.html](http://www.cpc.unc.edu/measure/prh/rh_indicators/overview/types-of-indicators.html), accessed 5 April 2016).
24. Aqil A, Lippeveld T, Hozumi D. Prism framework: paradigm shift for designing, strengthening and evaluating routine health information systems. *Health Policy Plan*. 2009;24(3):217–28. doi:10.1093/heapol/czp010.
25. Tools for data demand and use in the health sector: Performance of Routine Information Systems Management (PRISM) tools. Chapel Hill (NC): MEASURE Evaluation; 2011 (<http://www.cpc.unc.edu/measure/resources/publications/ms-11-46-d>, accessed 27 April 2016).
26. Weiss W, Tappis H. ePRISM: Adapting the PRISM framework for the evaluation of eHealth/mHealth initiatives. Baltimore (MD): Johns Hopkins University; 2011.
27. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE. Geneva: International Organization for Standardization (ISO); 2014 (ISO/IEC 25000:2014; [http://www.iso.org/iso/home/store/catalogue\\_ics/catalogue\\_detail\\_ics.htm?csnumber=64764](http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=64764), accessed 7 April 2016).
28. Al-Qutaish RE. Measuring the software product quality during the software development life-cycle: an International Organization for Standardization standards perspective. *J Computer Sci*. 2009;5(5):392–7.
29. Boehm B. A spiral model of software development and enhancement. *ACM SIGSOFT Software Engineering Notes*, 1986;11(4):14–24.
30. Labrique AB, Vasudevan L, Chang LW, Mehl GL. H<sub>2</sub>O for mHealth: more “y”, or “o” on the horizon? *Int J Med Inform*. 2013;82(5):467–9. doi:10.1016/j.ijmmedinf.2012.11.016.
31. m4RH Kenya: results from pilot study. FHI 360; undated (<http://www.fhi360.org/sites/default/files/media/documents/m4RH%20Kenya%20-%20Results%20from%20Pilot%20Study.pdf>, accessed 5 May 2016).
32. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. Geneva: World Health Organization; 2010 (<http://www.who.int/healthinfo/systems/monitoring/en/index.html>, accessed 7 April 2016).
33. Improving the evidence for mobile health. A.T. Kearney; undated ([https://www.atkearney.com/documents/10192/687794/GSMA+Evidence+for+Mobile+Health\\_v20\\_PRINT+HIRES.pdf/691847bc-3a65-4710-87eb-cc4753e8afce](https://www.atkearney.com/documents/10192/687794/GSMA+Evidence+for+Mobile+Health_v20_PRINT+HIRES.pdf/691847bc-3a65-4710-87eb-cc4753e8afce), accessed 7 April 2016).
34. Palmer S, Torgerson DJ. Definitions of efficiency. *BMJ*. 1999;318(7191):1136.
35. Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988;260(12):1743–8. doi:10.1001/jama.1988.03410120089033.



## Chapter 3: Monitoring digital health interventions



## HOW WILL THIS CHAPTER HELP ME?

This chapter will:

- ✓ Identify key components for monitoring the inputs of digital health interventions, including defining what should be monitored by whom, and at which point(s) during the implementation process.
- ✓ Detail mechanisms to monitor the digital health interventions to ensure that implementation factors do not threaten the potential effectiveness of the intervention.
- ✓ Demonstrate how to use monitoring findings to make corrective actions and optimize implementation, which can in turn improve the success of evaluation efforts.

**T**he separate concepts of monitoring and evaluation – together known as M&E – can be difficult to disentangle. Both sets of activities are frequently conducted in parallel and presented as a linked pair. This chapter focuses on monitoring – particularly the monitoring of technical, user and programmatic inputs, also referred to as process monitoring. An extensive body of literature already exists on process monitoring; this chapter is therefore not a replacement for, but rather a supplement to this literature, with special consideration of the monitoring challenges and opportunities introduced during digital health interventions. By conducting adequate monitoring of digital health interventions, project managers can better ensure that technical system implementation does not threaten overall project effectiveness. Failure in digital health monitoring can lead to intervention failure. For example, if the intervention relies on successful sending and receipt of SMS messages but implementation teams do not regularly monitor SMS failure rates at the server, they would not find out that clients had not received messages until end-line surveys. This would result in missed opportunities to make prompt corrections to the system and prevent failure of the intervention. See Chapter 1 for more information on the distinctions between monitoring and evaluation.

Box 3.1 presents and defines the four major components of the monitoring of digital health interventions that will be used to guide this chapter: functionality, stability, fidelity and quality.

Figure 3.1 illustrates the interaction of these five major monitoring components (top half of the figure), along with shifts in the importance of each component for monitoring the intervention, based on changes in the stage of maturity over time. As shown, the focus shifts from monitoring stability and functionality during early stages of intervention maturity, to high-level monitoring of fidelity and quality as the intervention matures and grows towards national scale. Monitoring activities can be further broken down to address aspects of system quality, user proficiency and the fidelity with which a system and user – in tandem – consistently perform the stated or intended objectives. The lower half of Figure 3.1 shows the six major evaluation components, which will be discussed further in Chapter 4.



**Process monitoring:** The continuous process of collecting and analysing data to compare how well an intervention is being implemented against expected results (1). In this Guide (i.e. in the context of digital health interventions), “**monitoring**” and “**process monitoring**” are used interchangeably to refer to the routine collection, review and analysis of data, either generated by digital systems or purposively collected, which measure implementation fidelity and progress towards achieving intervention objectives.

**Monitoring burden:** The amount of effort and resources required to successfully monitor the intervention; this burden is driven by the stage of maturity, the size of the implementation, the amount of data, and the number of users and indicators to be monitored.

**Users:** The individuals who directly employ the technology using their mobile phones, either to deliver health services (e.g. community health workers, district managers, clinicians) or to receive services (i.e. clients, patients).

### Box 3.1. The four major components of digital health monitoring – definitions

**Functionality:** A “characteristic that represents the degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions” (2). In this Guide, functionality refers to the ability of the digital health system to support the desired intervention. Functionality may also be referred to as “functional suitability”.

*Answers the question: Does the system operate as intended?*

**Stability:** The likelihood that a technical system’s functions will not change or fail during use. In this Guide, stability refers to the ability of the digital health system to remain functional under both normal and anticipated peak conditions for data loads.

*Answers the question: Does the system consistently operate as intended?*

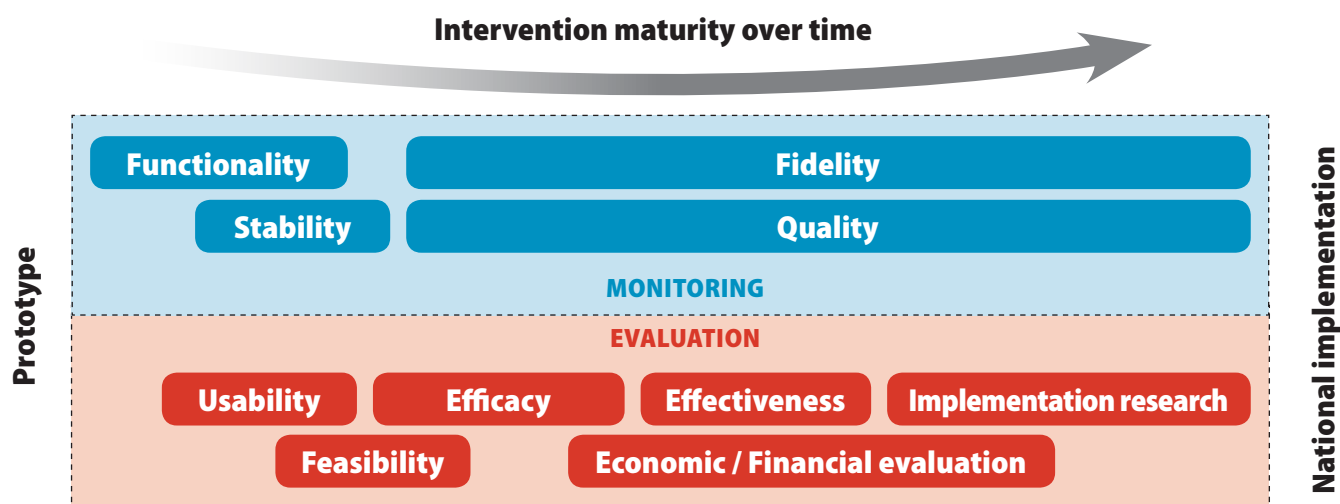
**Fidelity:** A measure of whether or not an intervention is delivered as intended (3). In this Guide, fidelity is viewed from both a technical and user perspectives.

*Answers the question: Do the realities of field implementation alter the functionality and stability of the system, changing the intervention from that which was intended?*

**Quality:** A measure of the excellence, value, conformance to specifications, conformance to requirements, fitness for purpose, and ability to meet or exceed expectations (4). In this Guide, the quality of a digital health intervention is viewed from both user and intervention content perspectives.

*Answers the question: Is the content and the delivery of the intervention of high enough quality to yield intended outcomes?*

**Figure 3.1. Intervention maturity life-cycle schematic, illustrating concurrent monitoring (blue/upper) and evaluation (red/lower) activities that occur as an intervention matures over time (left to right) from a prototype application to national implementation**



## What to monitor versus evaluate – identifying target inputs

Some inputs may be easy to identify – for example, the number of working mobile phones deployed. Inputs relating to how people interact with the system can be more difficult to define, but it is important to do so. Identifying and distinguishing which system interactions should be classified as inputs, outputs or outcomes will guide the management team in the selection and measurement of input indicators to be included in monitoring activities.

Start this process by asking, Is the user a primary user or secondary user? Primary users include the health workers or clients who directly interact with the digital health system. Secondary users are individuals who derive benefit from



primary end-users' input into the digital health system, but do not themselves directly enter data (e.g. supervisors or clients passively receiving text messages). The inputs measured in a given digital health intervention will differ based on the type of intervention it is, and these inputs can be categorized specifically by how targeted users or recipients interact with the system itself. Answering the question posed above is the first step in identifying the technical and user-linked inputs of the digital health intervention. The information in Box 3.2 can assist with making this determination.

### **Box 3.2. User interactions as primary versus secondary**

Primary users interacting with the digital health system . . .

- enter information on a digital health application and transmit the data
- rely on a digital-health-based algorithm to tell their clients if they are at risk for certain illnesses
- send SMS messages to learn more about contraception methods.

Secondary users interacting with the digital health system . . .

- receive SMS messages to remind them to take their medication or visit a health-care facility
- receive phone calls to educate them about hygienic practices for themselves and their families
- receive SMS messages to remind them to visit particular clients in a rural community.

This chapter will consider the first of these scenarios, where individuals, often community health workers or others on the supply side, are primary users of the digital health system. In these cases, technical and user inputs are difficult to disentangle while monitoring the fidelity and quality components of the intervention and so are presented in tandem. In some implementations, there may be both types of users.

This chapter focuses on monitoring, so (although there is interdependence between monitoring and evaluation) the emphasis here is placed on intervention inputs (i.e. what goes into implementing the digital health system, such as technical functionality and stability, and quality of implementation), rather than on outputs or outcomes (i.e. what comes out as a result of the implementation, such as 90% of immunization reminder messages being read by the target family within 1 day of delivery, or a 30% increase in polio vaccine coverage in a particular area). See Chapter 4 on Evaluation for more details on how to evaluate the output and outcome indicators specific to digital health system interactions. See the example box at the end of this chapter for more on how to identify user-linked project inputs.

Like any other intervention, implementers of digital health systems will want to ensure that inputs are of the highest quality possible. In non-digital health interventions, this means providing high-quality training, theory-based practices for behaviour change communication messaging (5), or high-quality supplements with the expected nutrient content (6). Similarly, in the case of digital health projects, ensuring high-quality inputs – such as training that is adequate to support high-quality worker performance, or sanitation and hand-washing message content that is in line with strategies known to be effective – is key to ensuring the eventual effectiveness of the digital health intervention (7). Unlike non-digital health interventions, additional monitoring must be conducted to ensure that the digital health application itself, the support systems and the users meet the specified standards and are truly ready for deployment and sustained use.

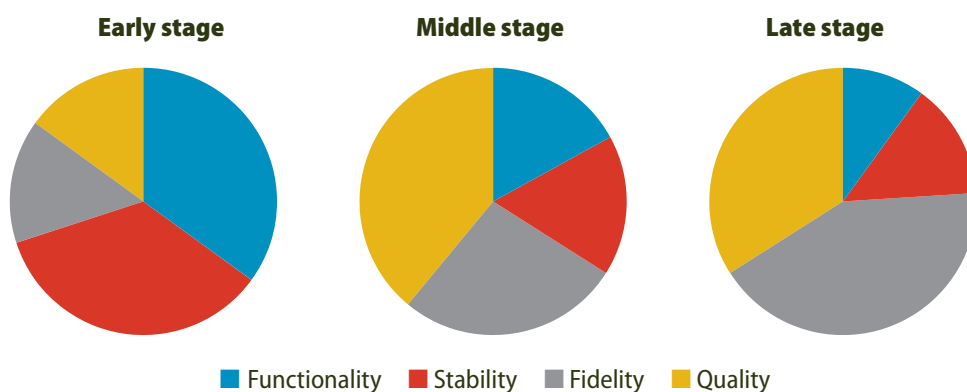
## Part 3a: Identifying stages of intervention maturity

Another important question when setting up a monitoring plan is: Which maturity stage is this intervention in? An honest and accurate answer to this question is critical because it will allow the project teams to budget time and resources accordingly. Projects in the early stages of maturity should consider investing significant resources – financial, human and time – to ensure that a newly created system is functional and stable. For projects at a more advanced stage of maturity, implementers might instead want to focus on the quality and performance monitoring components, dedicating resources not towards the technical aspects of the implementation but towards scale-up and continuous improvements to the system.

Interventions may also fall somewhere in between this continuum of maturity. An intervention at this middle stage of maturity has likely had to make significant upgrades to its pilot-tested technology to make it more robust and user-friendly, but the basic functionality, system structures and testing infrastructure already exist. With an increased number of users and an expanding geographical area, implementers will need to think through standardization of training, benchmarks for worker performance and a systematic approach to day-to-day monitoring of both the workforce and the technology; these are new considerations that did not need to exist during earlier stages of maturity.

Figure 3.2 illustrates the relative monitoring burden by component that could be expected by projects in each stage of maturity. As shown in Figure 3.2 (and reflecting the information in the top half of Figure 3.1), interventions in the early stages have a higher burden of technical monitoring (functionality and stability) and those in later stages have a higher burden of implementation-linked monitoring (fidelity and quality).

**Figure 3.2. Relative monitoring burden by component across intervention maturity stages**





## Part 3b: Tools for monitoring

This Guide assumes that several critical steps in system development and monitoring preparation have been undertaken, including careful definition of the needs of the users, and an engaged, user-centred design process, as well as data completeness checking, among others. These “raw materials” are useful in setting up a monitoring plan as well. For more information on these processes, there are numerous tools available, a few of which are listed in Box 3.3 and briefly described below.

### Box 3.3. Raw materials checklist

- |   |                   |
|---|-------------------|
| ✓ Human-centered design (HCD)               | ✓ Data            |
| ✓ Software requirements specification (SRS) | ✓ Codebooks       |
| ✓ Use case narratives                       | ✓ Indicators list |
| ✓ Wireframes                                | ✓ Dashboards      |
| ✓ Quality assurance (QA) test cases         |                   |

**Human-centered design (HCD):** HCD is a process in which the needs, wants and limitations of end-users of a product are given extensive attention during its design and development (8). Also referred to as “user-centred design”, designing digital health systems with the users (both on the delivery and receiving end) in mind is key to developing successful systems (9), by improving the quality and fidelity of the programme. Intervention quality improves when the system is made easier for users to operate in the way they are trained, and when content is being delivered in a more easily understood way that can be acted on by recipients. Fidelity may be improved by promptly addressing unanticipated external barriers that affect system usage. This useful toolkit from IDEO on HCD may help to guide your team through the process: <https://www.ideo.com/work/human-centered-design-toolkit/> (10).

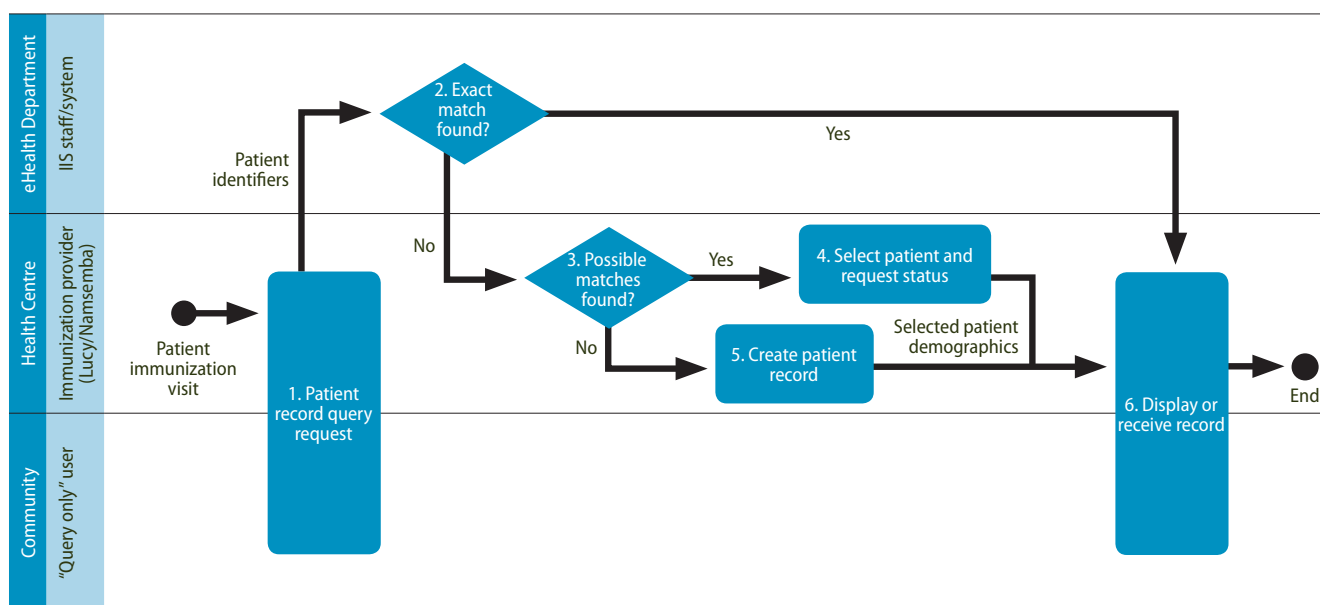
**Software requirements specification (SRS):** The SRS is a document that outlines the technical requirements of a desired system, clearly outlining requirements from the team of health experts and project managers for the developers responsible for creating the project’s digital health system. The SRS, also sometimes referred to as “business requirements”, serves as a touchstone document throughout development and the iterative communication/development processes. Process flow diagrams (or business flow) may also be incorporated into the SRS to map how the system should function. Importantly, this document outlines the expected functionality of the system and helps developers understand the stability requirements early in the development cycle (see Figure 3.3).

A robust SRS, while it requires investment of resources up front, may save a project significant amounts of time in monitoring the functionality and stability. During system testing, the SRS serves as a quality assurance checklist for all the features that need to be developed and need to function as specified. The SRS should outline findings from the formative research in clear and concise language and figures that can be easily referenced at a later time. Standard approaches for investigating and documenting SRS should be leveraged. The collaborative requirements development methodology is an approach that is specific to ICT for public health approaches (11). This SRS outline from Dalhousie University using IEEE standards is also valuable: [https://web.cs.dal.ca/~hawkey/3130/srs\\_template-ieee.doc](https://web.cs.dal.ca/~hawkey/3130/srs_template-ieee.doc) (12).

**Use cases:** Use cases are defined as narrative descriptions of how a target user performs a specific task using the technology and how the system is expected to respond to each case (13). Use cases are often included as part of the SRS. The information at this link provides a helpful guide on how to write successful use cases: <http://www.usability.gov/how-to-and-tools/methods/use-cases.html>.

**Wireframes:** Wireframes are simple, schematic illustrations of the content, layout, functions and behaviour of the target system (14); they are useful in illustrating the expected functionality of a system. For suggestions and guidelines on creating wireframes, see: <http://www.usability.gov/how-to-and-tools/methods/wireframing.html>.

**Figure 3.3. Illustrative examples of business process flows and SRS**



ID	ACTIVITY	REQUIREMENT (THE SYSTEM MUST OR SHOULD . . .)
10.5	Identify groups of vaccination events for evaluation	Allow users to manually flag duplicate events
10.6	Identify groups of vaccination events for evaluation	Have ability to display to the end-user the vaccine type, manufacturer, administrator date and eligibility
10.7	Identify groups of vaccination events for evaluation	Support a rules-based algorithm to evaluate duplicate events
10.8	Identify groups of vaccination events for evaluation	Support probabilistic algorithm to determine and flag when duplicate events need manual review
10.9	Identify groups of vaccination events for evaluation	Allow rules to be easily editable by staff when authorized
10.10	Duplicate events	Allow user to manually flag events for manual review
10.11	Duplicate events	Have ability to alert user of events pending for manual review
10.12	Duplicate events	Allow users to view events and event details simultaneously for decision merge
10.13	Duplicate events	Allow user to navigate the system while reviewing possible duplicates
10.14	Select the most accurate event record	Have ability to automatically select the most accurate/suitable vaccination event to be used as the primary record
10.15	Update vaccine event records	Allow user to select data elements to merge into a consolidated event record
10.16	Update vaccine event records	Have ability to combine two or more duplicate event records according to business rules
10.17	Update vaccine event records	Support an audit trail when event records are merged
10.18	Update vaccine event records	Have ability to retain "pre-merged" event records
10.19	Update vaccine event records	Have ability to generate an audit list of vaccination events that are automatically merged
10.20	Update vaccine event records	Allow user to delete a duplicate vaccine event while still maintaining audit record

Source: PATH, 2014 (15).

**Quality assurance (QA) test cases:** QA test cases are short sentences or paragraphs that describe expected functionality of discrete system functions and the steps to follow to perform each function. QA test cases break the more narrative or graphical functionality descriptions from the use cases and SRS into single-statement functions and expected actions. Using these test cases, implementation teams can test if the expected action actually occurs and if there is any deviation from what is expected or required. The QA test cases therefore facilitate a systematic process by which to guide and record feedback during complicated system testing.

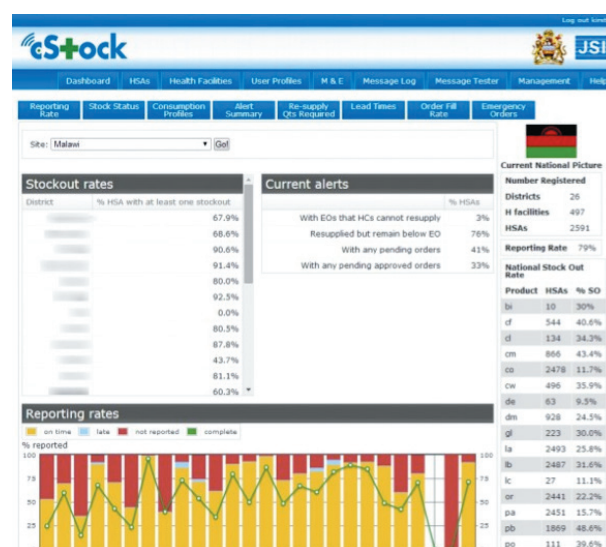
**Data:** A substantial amount of work goes into defining the technical specifications for any digital system. With so many moving pieces and looming deadlines, front-end, user-centred interfaces are often prioritized while back-end data systems get lost in the shuffle. As early as possible in the development process, or at least before implementation, the programme management team should check in with technologists to ensure that the data they will need for monitoring and evaluating implementation will be available, complete and usable. Care should be taken not to simply assume that data points which are critical to monitoring or reporting from a project manager’s perspective are being collected in the back-end; technologists may not have the same view of which data points are important. Programme managers may be interested in looking at how many immunization events did *not* occur – a non-event that is critical for tracking immunization worker accountability. In thinking through which data are required, the team should take advantage of the possibilities afforded by mobile data collection, such as timestamps or GPS locations to track when, where and for how long users send data. The data structure should be double-checked as soon as possible to avoid mistakes or missed opportunities.

**Codebooks:** Codebooks, also known as “data dictionaries”, provide a description of a data set that details features such as the meaning, relationships to other data, origin, usage and format of specific data elements (16). Before actually being able to analyse any data, or even use it for monitoring a programme, the definitions of each data point must be clearly communicated between the development and programme teams. Depending on the project, the codebook may be generated by the development teams after development is completed, or preferably by the programme team before development begins – to ensure that all the necessary data will be collected, accessible and analysable.

**Indicators list:** What are the key indicators that will be monitored and evaluated during this programme? How frequently will they be monitored and reported on? The programme team should do a final check on the data sources for each indicator to ensure that what needs to be measured can be – if possible do a dry run using test data to ensure all the pieces are in place before going live. An indicators list, with action items for when and how each point will be assessed, becomes the roadmap that guides the monitoring plan. This should incorporate input indicators addressing each of the monitoring components discussed in this chapter (see Part 2c: Indicators for more information on how to develop SMART indicators).

**Data dashboard:** Data dashboards are user interfaces that organize and present information and data in a way that facilitates interpretation (8). Access to more data more frequently does not necessarily translate into better monitoring, unless there is time for meaningful data review and data-driven decision-making. Depending on the size and scope of the project, development of basic dashboards – displaying high-priority indicators by time point, worker, location, or summary overview of key data needs – reduces the burden of report generation that tends to slow down analysis of data. Dashboards can also use visualizations to help project managers understand at a glance how workers or systems are performing, which may not be immediately apparent when presented in tabular format. Figure 3.4 is a screenshot from the cStock programme’s dashboard, which uses real-time data to provide managers with the tools they need to make data-driven decisions in real time, including “alerts, stock-out rates and current stock status” (17).

**Figure 3.4. cStock’s monitoring dashboard**



## Part 3c: Digital health process monitoring components

This section will examine in detail each of the four major components for digital health process monitoring that are defined in Box 3.1 and summarized in Table 3.1: functionality, stability, fidelity and quality. Here, for each component in turn, we will look at what to monitor, how to monitor, who will monitor, when to monitor, how to use monitoring findings, and how to monitor differently by maturity stage.

### Components for process monitoring

A summary of each monitoring component is presented in Table 3.1, including the primary objective of conducting monitoring within that component (presented as an overarching descriptive question), when it should be monitored in relation to intervention launch, examples of the potential inputs, the aspects on which that component focuses (technical, user interaction or implementation), and the burden by stage of maturity (also illustrated in Figure 3.2).

**Table 3.1. Summary of process monitoring components**

Component	When	Guiding question	Potential measures	Category	Monitoring burden by maturity stage
<b>Functionality</b>	Pre-launch	Does the system operate as intended?	<ul style="list-style-type: none"> <li>SMS content</li> <li>SMS schedules</li> <li>SMS timing</li> <li>Form content</li> <li>Form schedules</li> <li>Application functions</li> <li>Comparison of requested system vs delivered system</li> <li>QA test case adherence</li> </ul>	Technical	Early: High Mid: High Late: Low
<b>Stability</b>	Pre-launch	Does the system consistently operate as intended?	<ul style="list-style-type: none"> <li>Server downtime</li> <li>SMS failure rate</li> <li>Network connectivity</li> <li>Server operation capacity</li> </ul>	Technical	Early: High Mid: Medium Late: Low
<b>Fidelity</b>	During implementation	Do the realities of field implementation alter the functionality and stability of the system, changing the intervention from that which was intended?	<ul style="list-style-type: none"> <li>Stability reports</li> <li>Functionality reports</li> <li>Phone loss or damage</li> <li>Poor network connectivity</li> <li>Power outages</li> <li>User forgets password</li> <li>Incorrect intervention delivery by user</li> </ul>	Technical + user interaction	<b>Technical</b> Early: High Mid: High Late: Low  <b>User</b> Early: High Mid: Low Late: Low
<b>Quality</b>	Pre-launch & during implementation	Is the content and the delivery of the intervention of high enough quality to yield intended outcomes?  How well and consistently are the users delivering the intervention?	<ul style="list-style-type: none"> <li>User entry of phone number is correct</li> <li>Rate of agreement in data recording between training rounds (i.e. user accuracy)</li> <li>Quality control reports on users</li> <li>Feedback from users on content</li> <li>Incorrect schedules or content updates</li> <li>Timestamps on form submissions</li> <li>Number of form submissions/worker</li> <li>Data patterns similar across workers/geographic areas</li> </ul>	User interaction + implementation	<b>User</b> Early: Low Mid: High Late: High  <b>Content</b> Early: High Mid: Low Late: Low

## Functionality

In this section we discuss how to assess the functionality of technical systems.

Before launching any digital health system, extensive testing should first be carried out to ensure that the system is operating as intended and is free of bugs. A logical place to start is by defining what it means to be “operating as intended”. If the system has an SMS application, use this as the starting point to create a guided testing document, also referred to as QA test cases (see Part 3b). Based on findings from usage of these QA test cases, an iterative process of feedback to developers, additional development and re-testing will likely be necessary before arriving at a “final” or field-ready system that contains the necessary ingredients to deliver the intended intervention. Both front-end (user) and back-end (data and process) systems need to be tested to ensure adequate system functionality.



**What to monitor:** Depending on the type of application and system that has been developed, first consider testing and providing feedback on skip patterns, validation checks, form schedules, form content, user interface (UI) design, data export functionality, data accuracy and dashboard calculations. Flow diagrams developed for the SRS will be useful in testing skip patterns, validation checks and form schedules, while mock-ups of application interfaces can be useful in providing feedback on UIs and in-application functionality and flow. The key questions to ask are:

- Does the system meet the requirements outlined in the SRS?
- Does the system meet the needs of the health intervention?



**How to monitor:** As shown using an example in Table 3.2, QA test cases can help coordinate the testing process between developers, project managers and field staff, outlining what is expected to occur (e.g. “New Woman Registration Form v1.0” is launched) when the user does a specific action (e.g. user clicks “Add new woman” button), and systematically recording the test case’s status (pass or fail) on whether or not the expected outcome actually occurs (e.g. Fail: user clicks “Add new woman” button and system launches “New Child Form v2.0”). Creating these QA test cases in advance for all functions of the system or application helps ensure that no blocks of functionality are accidentally left out during this important testing phase.

**Table 3.2. Example QA test cases for two functions**

Test case	Scenario	Expected output	Actual output	Status
New woman (client) is found by health worker	Health worker user clicks “Add new woman” button	New Woman Registration Form (v1.0) is launched	New Child Form (v2.0) is launched	Fail
Polio-1 vaccine given	Health worker user clicks “Administered Polio-1 vaccine”	<ul style="list-style-type: none"><li>■ Polio-1 vaccine displays as “given” with date given</li><li>■ Polio-2 vaccine is scheduled at polio-1 date + 4 weeks</li></ul>	<ul style="list-style-type: none"><li>■ Polio-1 vaccine displays as “given” with date given</li><li>■ Polio-2 vaccine is scheduled at polio-1 date + 4 weeks</li></ul>	Pass



**Who will monitor:** Successful functionality monitoring will depend on having the human resources available to assign to this task, and this will be partially dictated by the intervention’s stage of maturity. In early stages, the project manager may conduct the bulk of this monitoring, whereas in later maturity stages he or she may be able to delegate this task to other staff members who are familiar with the expected functionality of the system and comfortable using the QA test cases. Individuals with a strong field presence, such as field supervisors, may test the content of the intervention for accuracy, including skip patterns and logic, SMS content or schedules.



**When to monitor:** The first push towards system finalization, comprising iterative feedback and development loops between the testing team and developers, should be completed before the launch, always keeping in mind that it is usually easier to make changes to a digital health system before it goes live. Continued functionality monitoring, under the umbrella of fidelity monitoring, should continue even after a system is deemed functional and launched, especially during the first weeks and months of implementation, as problems may arise in real-world usage that didn’t surface during desk-based or preliminary testing.



**How to use monitoring findings:** All pre-launch testing findings should be compiled regularly and shared with developers to promptly resolve any identified problems. This is particularly important during continued functionality monitoring after launch, as functionality shortfalls are then affecting live intervention services.

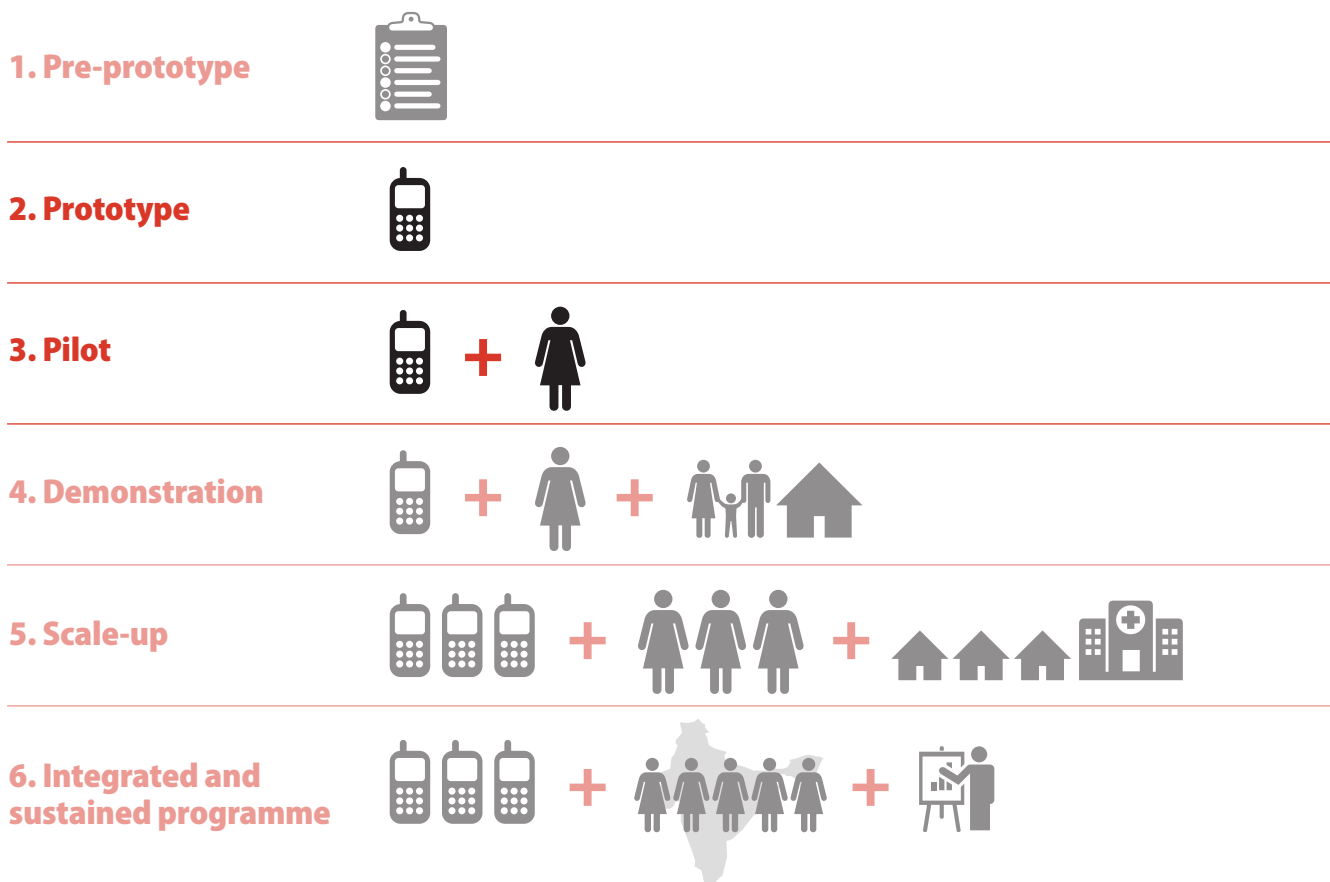
Any intervention services that are disrupted by functionality problems should be well documented and taken into consideration during the evaluation phase. For example, if a technical problem that prevented SMS vaccination reminders persisted in the system for 15 days, and during that time 110 families missed the opportunity to receive the reminder messages, then later evaluation analysis of the impact of the intervention may need to take into account that this number of families in the target population were not exposed to this particular component of the intervention (vaccination reminder messaging).



**How to monitor differently by maturity stage:** Pre-launch functionality monitoring is most important and most burdensome for early stage digital health systems that have never been implemented, or are being fielded for the first time under different circumstances from previous implementations.

These interventions in the early stages of maturity (see Figure 3.5) are also likely to have the highest burden of continued functionality monitoring requirements throughout the duration of the intervention. As consistent system functionality has not yet been demonstrated for the system, projects in early stages of maturity should allocate substantial resources to both pre-launch and continued functionality monitoring. Interventions in later stages of maturity should conduct basic functionality monitoring before re-launching when introducing the system to a different cadre of health workers (i.e. new users), new geographic areas that may pose different levels of connectivity, or when using new technologies. The interventions in more mature stages should also continue with monitoring efforts during implementation, but can focus most of these efforts and resources on the fidelity and quality components.

**Figure 3.5. Interventions in stages 2 and 3 of maturity (pilot-testing and limited demonstration) will need to focus on monitoring functionality and stability**





## Stability

In this section we discuss how to monitor the stability of technical systems.

Monitoring of system stability is semi-concurrent with functionality monitoring, but it brings additional monitoring requirements post-implementation. Unstable systems will perform unreliably, crash or stop unexpectedly, slow down when overloaded or otherwise perform erratically. Poor stability may result in improper delivery of the intervention. For example, the system may frequently fail to deliver vaccination reminder SMS messages or unreliable performance may make users hesitant to use the digital health intervention as intended. A key characteristic of stability monitoring is that it can be largely automated after initial testing and during launch.



**What to monitor:** Digital health applications or systems that rely on physical (non-cloud-based) servers for a portion of their operation may find that server outages are a primary source of instability. During pre-launch monitoring of stability, the cause of these outages should be identified to the furthest extent possible (e.g. power failure or data overload) and efforts made to minimize the risk of future outages.

- What is the failure rate of SMS messages from the server side?
- If there is a UI to the system, how often are there unexpected application closes, crashes or forced quits?
- How responsive is the digital health system under both normal and anticipated peak conditions for data loads?



**How to monitor:** Server logs can be used to identify the events that lead up to a server outage. When possible, as soon as an outage is detected this should trigger automatic alerts to server support teams, to increase the likelihood of diagnosing the problems that led to the outage. SMS client servers record success and failure statuses for all messages sent, so it is possible to monitor the failure rate and to set a cut-off point for when this rate is deemed unacceptable for the project.



**Who will monitor:** In many cases, the technical development team – the people who develop and maintain the system or server – will be responsible for collecting server logs or application crash reports and for diagnosing and reporting on the causes of outages and instability. The project manager should sit with this individual or team to understand the source of these problems and if there is anything that can be done to prevent repeat outages in the future that cannot be done from the technology development team's side. Additional work may be required by the technology development team to reduce the likelihood of a similar crash in the future, such as optimizing the application so that it runs more efficiently.



**When to monitor:** As with functionality monitoring, a first round of stability monitoring should be conducted well in advance of intervention launch. Unlike functionality monitoring, however, it may be difficult to get a full picture of system stability during the testing phase. For example, if the project has 500 system users, there may never be an opportunity to test form submissions, or other similar measures in the volume that will occur once the project goes live and has been running and accumulating these events over the course of months or years. Setting up systems for continuous stability monitoring is critical for the duration of the intervention, and is part of continued stability monitoring under the umbrella of fidelity monitoring in later stages of programme maturity.



**How to use monitoring findings:** Despite extensive pre-testing and other precautionary measures, issues will inevitably arise. Having automated systems in place to monitor stability is feasible at various maturity stages, particularly at the server level. As the intervention moves towards later stages of maturity, these systems will need to become more sophisticated. To decrease downtime, alert messages to server managers should be triggered when systems go down or automated code can be set up to manage server usage before failure occurs. Data on system downtime should be reviewed to look for patterns of instability that can be used to resolve the problems.



**How to monitor differently by maturity stage:** Stability monitoring is most important during the pre-launch phase of a project, but it remains a high priority throughout implementation for interventions in early and later stages of maturity. For interventions in later stages of maturity, automated systems can be developed to track system stability and immediately inform project managers and supervisors of any instability detected. Investing resources in robust, automated stability-monitoring features should reduce the amount

of downtime experienced by the large number of system users of digital health interventions in later stages of maturity. Importantly, as implementations expand, so too must the technical systems that support them; project managers must be careful to test expanded systems for functionality even if previous versions of a system were fully functional, since scaling systems from 1000 to 10 000 users may have huge effects on system stability.

## Fidelity

After the functionality and stability of the digital health intervention have been initially assessed and found to be adequate for system launch and early implementation, project managers should shift their approach to continued monitoring of both of these components for the duration of the project. At this point in implementation, however, it is not just the technical system that must be monitored. Other questions include:

- Are users using the application appropriately throughout the intervention period, to ensure the greatest possible value can be derived from the digital intervention?
- Are there any barriers to high fidelity intervention implementation (i.e. is there any reason, aside from technical functionality and user capacity, that could prevent intervention delivery)?

Monitoring fidelity can be divided into three broad categories: (a) monitoring the overall technical fidelity of the digital health system throughout the implementation process (i.e. assessing whether or not the system maintains stability and functionality throughout the duration of the intervention); (b) monitoring any barriers external to the defined system itself that are causing it not to function as expected (i.e. assessing if there are hardware issues or connectivity issues affecting the geographic area); and (c) monitoring compliance of digital health system users who mediate delivery of the intervention (i.e. assessing data on surveillance forms to ensure they are completed accurately).



### What to monitor:

*a. Technical* – Monitoring for errors and system stability does not end after the initial testing phase. Even after extensive testing, systems may not function as expected once they “go live” in their intended field settings. Errors and instability may occur for a number of reasons, such as poor network connectivity in the most rural regions of the deployment area, or having 600 field workers sending data simultaneously and overwhelming server capacity (continued stability monitoring), or previously functioning SMS messaging may malfunction after a new version is released (continued functionality monitoring).

Identifying key inputs of system performance to monitor before launch will help project teams take advantage of the real-time monitoring capabilities digital health systems can offer and resolve issues as quickly as possible. Specific components to consider monitoring include:

- Does the server experience uptime interruptions?
- What are the average SMS failure rates?
- How many forms are reported as sent versus actually received?
- What is the average time for form completion, amount of data usage, and number and timing of form submissions?

*b. External* – There are a range of external contingencies that are required for intervention delivery. Some external issues to consider are the following:

- What are the supportive materials required for consistent delivery of the intervention? (e.g. power banks to ensure the digital device is always charged)
- Do all health workers have access to the updated application that supports the updated content?
- Is the tablet used for data collection functional, charged and not lost or stolen?
- Do all health workers have enough data left in data subscription packages to send data when required? Are there enough credits in the SMS server for messaging?

c. *User* – User fidelity refers to the digital health users’ consistent adherence to implementation protocols, both in how users interact with the digital health system (e.g. the time it takes for a worker to submit a pregnancy surveillance form) and their compliance with non-digital health-related training procedures (e.g. estimating the gestational age when interacting with a pregnant woman). Some user adherence questions to consider are the following:

- Are health workers sending in data collection forms as frequently as expected?
- Are health workers able to operate the digital health application as intended, outside the context of their training?
- Are health workers following the appropriate health protocols when conducting their work?



**How to monitor:** Systems should be set up for continuous, automated server uptime monitoring to ensure system stability, including alerts that should be set to trigger notifications when server capacity has almost been reached (before the system goes down) or emergency alerts that trigger notifications once the system does go down, which should include information on power outages or memory storage limits. In addition, using real-time data to monitor digital health users, those who are performing poorly or below cut-off levels can be brought in for strategic retraining on the technical system (e.g. how to smoothly complete and submit an interview form) or on the intervention content (e.g. how to identify a new pregnancy in the community).



**Who will monitor:** The day-to-day monitoring of fidelity of implementation will often be carried out by field-level supervisory staff, who will report adverse events (e.g. phone loss) and mitigating actions (e.g. replaced phone) to the project manager. Once data have been entered, whether metadata from timestamps or details from paper-based forms recording last menstrual period (LMP), the project manager is responsible for regular review of monitoring data to check on programme implementation. The project manager should have previously identified key user indicators of high-fidelity implementation that must be monitored during this stage. For example, if early visits to a mother and newborn are critical to the success of the intervention (i.e. delivery of early newborn care), the project manager should look at the time lag between date of birth and date of first visit to the mother and newborn. Users whose visits fall outside an acceptable window of time for these first visits should be interviewed to understand what is causing the delay and how it can be resolved.



**When to monitor:** Fidelity monitoring must occur throughout programme implementation. As programme implementation continues, the amount of effort required to conduct fidelity monitoring may decrease, as live technical issues, user issues and external issues are identified and resolved.



**How to use monitoring findings:** Continuous technical monitoring allows for immediate reaction in the event that the digital health system is no longer supporting the intervention as intended. Promptly responding to technical disruptions or failures will reduce the amount of downtime experienced by all users. Monitoring reports generated at the user level can also point out systematic errors and weaknesses that may exist in implementation of the intervention related to inappropriate use of the technology, poor worker motivation and lack of use, or misunderstanding of intervention content being delivered or related best practices. The greatest benefit of user-based fidelity monitoring is that it enables project managers to target specific users who need to be brought in for retraining or counselling.



**How to monitor differently by maturity stage:** Monitoring of user fidelity is important at all stages of maturity, but it becomes increasingly important to have standard monitoring procedures and reporting mechanisms in place as the number of users increases. In an early-stage pilot-test with 10 users, field supervisors can likely get a reasonable understanding of user strengths, weaknesses and usage by conducting field visits and through cursory reviews of incoming data. In programmes with large numbers of users – 600 users, for example – it will no longer be possible for a small number of supervisors to compile and act on monitoring findings. Displaying statistical performance data for too many workers is difficult to digest in tabular format and will likely require development of or plug-in to graphical software or even visual geographic systems through dashboard interfaces. Therefore, as the intervention moves through the stages of maturity, the degree to which monitoring is automated and reported for decision-making must advance as well.

# Quality

In this section we discuss two aspects that should be addressed while monitoring the quality of digital health interventions: (a) user capabilities and (b) intervention content.

- a. *User capabilities* – Training large numbers of workers on digital health systems presents many challenges – from low technical literacy and lack of previous exposure to smartphones, tablets or computers, to unwillingness to change work processes that have been in place for decades. Even within the same workforce, there may be wide variation between individual workers in their comfort with the hardware and software, depending on previous off-the-job experience with technology, which is often age-related, with younger workers starting out much more tech literate than their more senior counterparts (18). Key questions that need be answered include:
- Are the users (e.g. health workers) entering information accurately into the digital health system? This question points towards the readiness of the workers to begin using, or in some cases continue using, the system.
  - Are there gaps in user understanding that prevent correct system use or intervention delivery?
- b. *Intervention content* – This second aspect of quality monitoring relates to the quality of the content or intervention that the digital health system is trying to deliver. In other words, the content of the inputs (e.g. SMS messages, algorithms for decision support, data collection forms) used for the intervention should be of the highest quality possible, informed by the existing literature and formative research in the local context, to increase the effectiveness of the digital health intervention (7).



## What to monitor:

- a. *User* – Project managers need to determine what the key functionalities are for users of the system and what indicators can be used to determine user readiness. These indicators may measure key phone operations, knowledge of system functionalities, accuracy of data collection, or basic troubleshooting skills. Once the system is launched, quality monitoring can focus more specifically on data accuracy and regularity by checking for outliers of non-compliant users.
- Are all workers providing services in a similar way, in terms of the length, content and quality?
  - Are some health workers able to manage the cases of more clients than other workers?
  - Are there irregularities in the time or place from which the data are being sent by each health worker (checked through timestamps and GPS codes)?
  - Are there unusual patterns in data collection, such as follow-up visits being recorded consecutively within a short time span?
- b. *Intervention* – Before launch, it will be important to confirm that the content to be delivered is as expected and in line with existing international standards as well as being appropriate for the community where the intervention will be implemented.



## How to monitor:

- a. *User* – In a hands-on training session, the key knowledge indicators can be included on a checklist that all users (i.e. health workers) must “pass” before graduating to a group that is deemed ready to go live. In this way, implementers can ensure a baseline level of system proficiency among users before launch, hopefully limiting user-linked errors that prevent intended tasks from being completed consistently. Figure 3.6 provides an example of a trainee competency checklist from KEMRI’s TextIT project, which includes key smartphone and application operation functionalities. Another method for monitoring data accuracy is to issue standard scenarios to trainees to reinforce concepts from training, either in a role-play setting or narrative form. By comparing trainee results with the expected results, managers can assess the percentage agreement (i.e. accurate responses) per question and per trainee.

After launch, continued monitoring should look for outliers or underperformers in relation to key indicators. For example, a project may select “ability to correctly enter gestational age for pregnant women” as a key indicator for success and flag workers who identify pregnancies outside of an acceptable range of accuracy.

- b. *Intervention* – The project team should ensure that content is based on evidence-based guidelines (e.g. WHO guidelines) or existing ministry of health documentation. Additionally, the team should conduct focus group discussions and/or interviews in the community to tailor appropriate messaging. See Maar et al. (2016) for an example of how this process was used to create novel SMS messaging based on existing evidence-based health information (7). Before implementation, the tone and construct of messaging should be monitored for quality and acceptability to the target audience (19).



### Who will monitor:

- a. *User* – Trainers are often best placed to assess core competencies using the checklist method, but trainers may be biased and tend to pass their own trainees, so it is advisable to use more senior staff or trainers from different groups for this task.

- b. *Intervention* – Project managers may be responsible for the overall content of the intervention, but senior-level team members, such as principal investigators in research studies or project area officers, may weigh in on the content of the intervention to be delivered and are ultimately responsible for the quality of this content as it makes up the primary substance of the intervention.



### When to monitor:

- a. *User* – Monitoring users' comfort in interacting with a system is important in determining user readiness before the launch. Regular assessments should also continue throughout the duration of the intervention to ensure that users continue to operate the system as intended.

- b. *Intervention* – Monitoring of the content quality will likely occur in the early stages of an intervention, before the launch. In some cases, improvements may be made to the quality of this content based on user feedback. In these instances, additional quality monitoring should be conducted to ensure that the updated content fulfils the original aims of the intervention and maintains a high overall level of quality.



### How to use monitoring findings:

- a. *User* – The results of monitoring user quality will allow trainers and supervisors to (i) gauge whether particular workers need more training, (ii) target specific areas for retraining, (iii) bring all users up to a baseline level of quality before launch, and (iv) maintain a high level of quality after launch. Utilizing the real-time data provided by digital health systems enables this feedback loop to be much faster than was possible under traditional systems. With an efficient monitoring and feedback cycle, course-corrections in intervention delivery can occur almost in real time.

- b. *Intervention* – If content quality is poor or unlikely to have the desired effect on the target population, adjustments can be made to improve this incredibly important aspect of the intervention. In the case that content is found to be poor quality and implementation has already begun, project managers must prioritize content adjustments and continuous content monitoring until content is determined to be of sufficient quality to drive an effective intervention.



### How to monitor differently by maturity stage:

- a. *User* – Monitoring user quality is important at all stages of intervention maturity but becomes increasingly challenging and resource intensive as the number of users and geographic coverage increases. Early-stage interventions with 10–15 users can be easily monitored qualitatively by a small number of

**Figure 3.6. Quality control checks on user ability to operate digital health system**

Skills / Knowledge	Rating	Comments
Sec 1: Basic Understanding of HIV/AIDS, PMCT, Safe Motherhood & Early Infant Care		
Understands basics of HIV		
Has good understanding of PMCT		
Has good understanding Safe Motherhood and Early Infant care		
Sec 2: Understanding of TextIT Study		
Has Understood the Aim, Objectives of TextIT Study		
Knows Study Interventions (text to her and call)		
Ability to identify eligible clients from ANC, PMCT, MCH/PCAR/PC		
Ability to Record Cases via SMS using the Study guide		
Ability collect Data using Smart phone		
Ability to do the first Data QA/QC		
Understands all the Source Documents		
Sec 3: Study SOPs		
Understands Content and Format of Study SOP		
Understands SOP for Preparation of Study Documents		
Understands the Follow-up Appointment SOP		
Understands Data Management Activities SOP		
Sec 4: Study Log		
Ability to complete the Screening Master Log		
Ability to complete the Follow-up Appointment Log		
Ability to complete the Intervention Log		
Ability to complete the Preparation of Study Documents Log		
Sec 5: Data Collection Using Smart phone		
Ability to switch the phone on and off		
Ability to identify the ODK Application		
Ability to access the electronic form in the ODK		
Can enter, edit, save & send Data using the ODK		
Sec 6: Other Responsibilities at work		
Ability to complete the Intervention Log		
Ability to on-the-spot support Groups for PMCT Clients		
Ability to refer and link PMCT clients appropriately		
Ensures safety and confidentiality of client records and Smart phones		
Trainer's signature: _____ Date: _____		
Trainer's signature: _____ Date: _____		

Source: Example from KEMRI TextIT project. KEMRI Internal Research Protocol (unpublished); personal communication (e-mail), Thomas Odeny, KEMRI, 14 January 2016.



supervisory staff, and issues that arise can be solved on a case-by-case or one-on-one basis. Once the number of users increases beyond what is easily manageable by a skilled supervisory team, the need to automate, standardize and visualize user quality monitoring increases dramatically.

- b. *Intervention* – Content will likely be created and tested during early-stage interventions (pilot-testing and small-scale implementation) and potentially refined during mid-maturity stage interventions. Most mature interventions will already have well defined and quality-tested content that has been optimized for scale-up.

#### **Example: Identifying user interactions as primary versus secondary, maturity stage, and priority monitoring components for a digital health intervention**

Consider the example of a pilot intervention that uses a digital health system to send SMS messages to remind families when their infants are due for immunizations, with the aim of increasing vaccination coverage rates. In addition to the SMS component of the intervention, there is a simple health worker interface that allows vaccinators to register clients and record immunizations.

- a. **Digital health user interactions as primary versus secondary:** In this example there are two types of users who interact with the system: the families and the health workers. The families are receiving SMS messages, so their interactions with the system will likely be measured as outputs (e.g. the number of SMS messages the family received), or outcomes (e.g. number of infants brought in for vaccination after their families received the message). The health workers' interactions with the system are different: they use a system interface to register infants and vaccinations – the information they enter will be used by the system to generate future SMS messages. As a user delivering health services and information, components of how a health worker interacts with the system are important inputs. Variables to monitor may include the accuracy of information entered by the health worker (e.g. the family's correct phone number), and use of the system by the health worker during the vaccination session.
- b. **Identifying the stage of maturity:** Recognizing the stage of maturity will allow the project manager to dedicate resources effectively. For the intervention in this example, an SMS system and digital health worker interface are being used – and both need substantial monitoring for **functionality** and **stability** before and during launch. Before implementation begins, consider the quality of the intervention content being delivered and the readiness of the system users. Does the reminder messaging contain the right information to encourage families to seek care?

As implementation begins, managers will want to increase the amount of focus they place on **fidelity** monitoring. Were the training courses provided sufficient to transfer the required knowledge to the health workers? Are the health workers correctly recording phone numbers so families can receive SMS messaging?

- c. **Monitoring burden:** Here, burden refers to the amount of effort and resources required to successfully monitor the intervention; this burden is driven by the stage of maturity, the size of the implementation, the amount of data, and the number of users and indicators to be monitored. Before implementing the intervention, the project manager must be sure that the digital health system functions properly – in this case, that the combined SMS messages and SMS scheduling is operating satisfactorily and that the health worker interface performs as expected. To conduct this functionality monitoring before launching the intervention, the project manager may create and then utilize quality assurance (QA) test cases – outlining exactly what should happen when certain actions are performed and then recording what actually does happen for each of these test cases. Issues to be checked would include whether or not the content of the SMS messages is correct, and whether or not the SMS messages are sent at the right time to the right phone numbers. In other words, the manager needs to determine whether, once the system is deployed, will the right family receive a reminder about their child's upcoming polio vaccination before it is due?

*(continued on next page)*

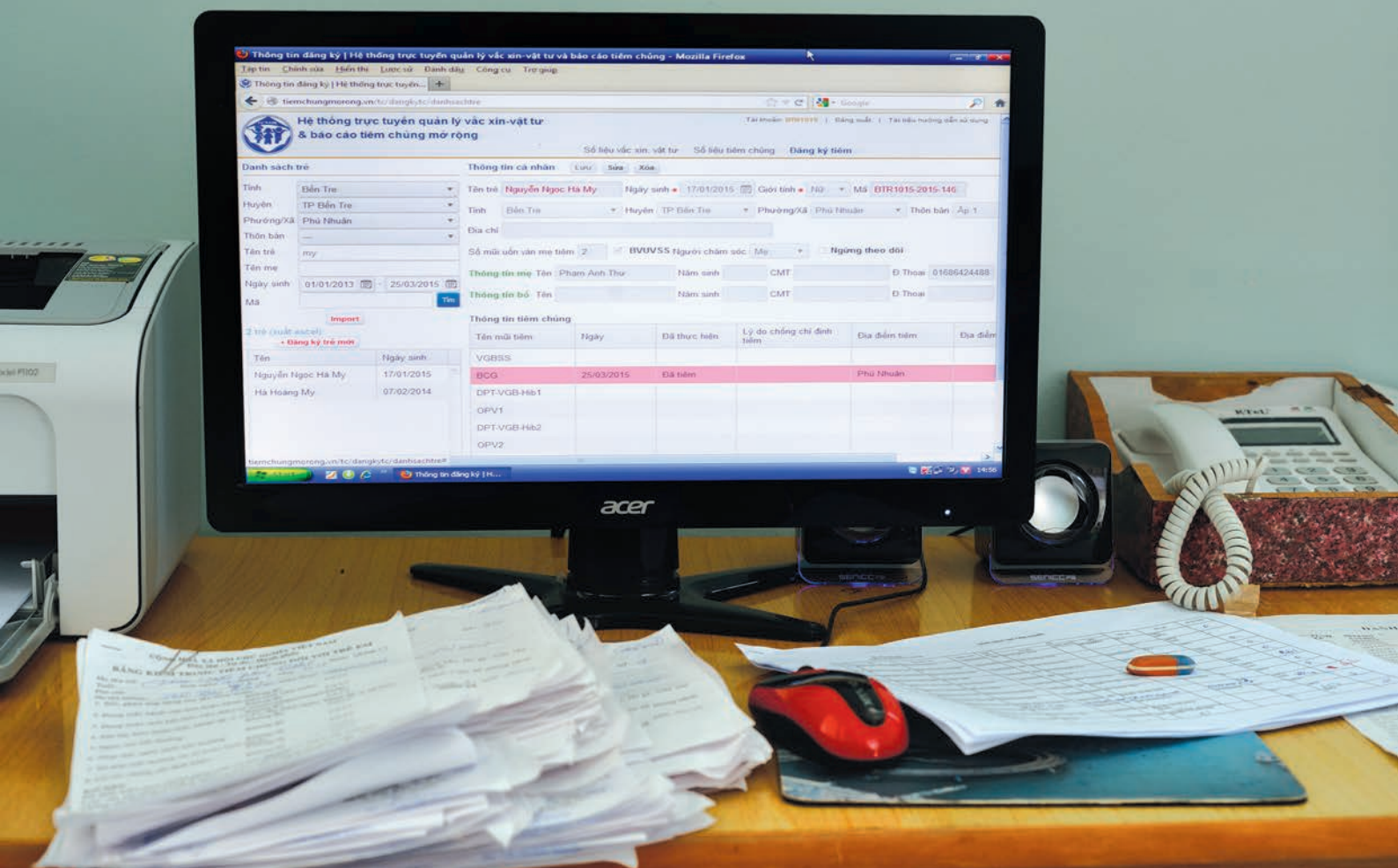


Once the basic functionality of the system has been assessed and deemed acceptable for the intervention, the project manager needs to verify that the intervention is stable. Many stability issues might be identified during initial functionality monitoring (e.g. frequent crashing of the health worker's digital application), but some stability issues may not be identified during the pre-launch period. To determine if the system is stable, the project manager and developers might check for downtime statistics of the server, SMS failure rates, and capacity for exceptionally high data burden that might overwhelm the system. For example, if there is a polio vaccination drive on Thursday, when all families in the area with children aged 0–5 years will receive an SMS reminder to attend the special session, can the SMS server handle this high volume of messages? In larger-scale interventions, stability monitoring can often be automated to report downtime or high capacity loads to the server team using alerts. This type of ongoing stability monitoring can help managers to course-correct during the early implementation period, helping to minimize incidents of technical failure that may prevent the intervention from being implemented as intended.

## References

1. WHO evaluation practice handbook. Geneva: World Health Organization; 2013 ([http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf), accessed 3 May 2016).
2. Functional suitability. In: ISO 25000 software product quality [website]. ISO 25000; 2015 (<http://iso25000.com/index.php/en/iso-25000-standards/iso-25010/58-functional-suitability>, accessed 9 May 2016).
3. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. *Implementation Science*. 2007;2:40.
4. Reeves CA, Bednar DA. Defining quality: alternatives and implications. *Acad Manag Rev*. 1994;19(3):419–45 (<http://www.jstor.org/stable/258934>, accessed 11 April 2016).
5. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? *Transl Behav Med*. 2011;1(1):53–71. doi:10.1007/s13142-011-0021-7.
6. West KP, Shamim AA, Mehra S, Labrique AB, Ali H, Shaikh S et al. Effect of maternal multiple micronutrient vs iron-folic acid supplementation on infant mortality and adverse birth outcomes in rural Bangladesh: the JiVitA-3 randomized trial. *JAMA*. 2014;312(24):2649–58. doi:10.1001/jama.2014.16819.
7. Maar MA, Yeates K, Toth Z, Barron M, Boesch L, Hua-Stewart D et al. Unpacking the black box: a formative research approach to the development of theory-driven, evidence-based, and culturally safe text messages in mobile health interventions. *JMIR mHealth uHealth*, 2016;4(1):e10. doi:10.2196/mhealth.4994.
8. Glossary. In: The mHealth planning guide: key considerations for integrating mobile technology into health programs. Toolkits by K4Health [website]; 2015 (<http://www.k4health.org/toolkits/mhealth-planning-guide/glossary>, accessed 28 April 2016).
9. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, Cafazzo J. mHealth consumer apps: the case for user-centered design. *Biomed Instrum Technol*. 2012;46(s2):49–56. doi:10.2345/0899-8205-46.s2.49.
10. Design kit: the field guide to human-centered design. IDEO; 2016. (<https://www.ideo.com/work/human-centered-design-toolkit/>, accessed 11 April 2016).
11. Common requirements for logistics management information systems. Seattle (WA): PATH; 2010 ([http://www.path.org/publications/files/TS\\_lmis\\_crdm.pdf](http://www.path.org/publications/files/TS_lmis_crdm.pdf), accessed 1 December 2016).
12. Wiegers KE. Software requirements specification for <project> (Template). Institute of Electrical and Electronics Engineers (IEEE); 1999 ([https://web.cs.dal.ca/~hawkey/3130/srs\\_template-ieee.doc](https://web.cs.dal.ca/~hawkey/3130/srs_template-ieee.doc), accessed 28 April 2016).
13. Use cases. In: usability.gov [website]. Washington (DC): United States Department of Health and Human Services; 2016 (<http://www.usability.gov/how-to-and-tools/methods/use-cases.html>, accessed 11 April 2016).

14. Wireframing. In: usability.gov [website]. Washington (DC): United States Department of Health and Human Services; 2016 (<http://www.usability.gov/how-to-and-tools/methods/wireframing.html>, accessed 11 April 2016).
15. Product vision for the Better Immunization Data (BID) Initiative. Seattle (WA): PATH; 2014 ([https://www.path.org/publications/files/VAD\\_bid\\_product\\_vision.pdf](https://www.path.org/publications/files/VAD_bid_product_vision.pdf), accessed 1 December 2016).
16. Hagan D, Uggowitz S (eSHIFT Partner Network). Information and communication technologies for women's and children's health: a planning workbook for multi-stakeholder action. Geneva: The Partnership for Maternal, Newborn & Child Health (PMNCH); 2014. ([http://www.who.int/pmnch/knowledge/publications/ict\\_mhealth.pdf](http://www.who.int/pmnch/knowledge/publications/ict_mhealth.pdf), accessed 11 April 2015).
17. mHealth cStock. Supply Chains 4 Community Case Management (SC4CCM); 2016 (<http://sc4ccm.jsi.com/emerging-lessons/cstock/>, accessed 4 May 2016).
18. Kaphle S, Chaturvedi S, Chaudhuri I, Krishnan R, Lesh N. Adoption and usage of mHealth technology on quality and experience of care provided by frontline workers: observations from rural India. *JMIR mHealth uHealth*. 2015;3(2):e61.
19. Pollard CM, Howat PA, Pratt IS, Boushey CJ, Delp EJ, Kerr DA. Preferred tone of nutrition text messages for young adults: focus group testing. *JMIR mHealth uHealth*. 2016;4(1):e1. doi:10.2196/mhealth.4764.



# Chapter 4: Evaluating digital health interventions



## HOW WILL THIS CHAPTER HELP ME?

This chapter will:

- ✓ Review basic concepts and key terminology in programme evaluations for digital health interventions.
- ✓ Introduce basic qualitative, quantitative and economic evaluation methods.
- ✓ Help you to define which evaluation activities are right for your digital health intervention.

Chapter 1, Part 1a introduced and described evaluation, and distinguished between monitoring (is the intervention doing things right?) and evaluation (is the intervention doing the right things?) (2).

Evaluation is optimally an ongoing cyclical process that informs adjustments and improvements to further intervention planning and implementation. Evaluation activities generate data that can be analysed and interpreted, forming evidence about the likely impact of the intervention.



**Evaluation:** The systematic and objective assessment of an ongoing or completed intervention, with the aim of determining the fulfilment of objectives, efficiency, effectiveness, impact and sustainability (1). In this Guide (i.e. in the context of digital health interventions), evaluation is used to refer to measures taken and analysis performed to assess (i) the interaction of users or a health system with the digital health intervention strategy, or (ii) changes attributable to the digital health intervention.

This chapter focuses on how to generate such evidence in the context of digital health interventions, for the purpose of evaluating their effectiveness, value for money and affordability. A central concept covered in this chapter is the framework for the different stages of evaluation that correspond to the various stages of maturity of the digital health intervention. The stages of evaluation, which are further elaborated later in this chapter, include the following:

- **Feasibility:** Assess whether the digital health system works as intended in a given context.
- **Usability:** Assess whether the digital health system is used as intended.
- **Efficacy:** Assess whether the digital health intervention achieves the intended results in a research (controlled) setting.
- **Effectiveness:** Assess whether the digital health intervention achieves the intended results in a non-research (uncontrolled) setting.
- **Implementation research:** Assess the uptake, institutionalization and sustainability of evidence-based digital health interventions in a given context, including policies and practices.

# Part 4a: Key concepts for conducting digital health evaluations

## Efficacy versus effectiveness: Can it work? Does it work?

To classify what stage a digital health intervention is at for the purposes of evaluation, you first need to consider the context in which implementation is occurring. In the context of an efficacy study, in which the intervention is delivered and received perfectly according to design under highly controlled conditions, evaluation will ask the question, Can it work? or What is the precise effect this strategy can have on my outcome, under ideal delivery and uptake conditions? (3). In the context of an effectiveness study, on the other hand, in which the intervention is implemented in a real-world setting such that delivery and response is not necessarily optimized (4), evaluation will ask the question, Does it work? (3) (see Box 4.1). A common approach in health systems research is to define effectiveness according to a continuum of outputs, outcomes and impact, as previously outlined in Chapter 2, Part 2b.

### Box 4.1. Efficacy versus effectiveness

**Efficacy** asks whether the intervention works in principle under ideal conditions.

**Effectiveness** asks whether the intervention actually works in a real-world setting.

Effectiveness can be assessed in terms of:

- **Outputs:** The direct products/deliverables of process activities in an intervention (5). From a digital health perspective, outputs can include improvements in performance and user adoption.
- **Outcomes:** The intermediate changes that emerge as a result of inputs and processes. Within digital health, these may be considered according to three levels: health systems, provider and client.
- **Impact:** The medium- to long-term effects produced by an intervention; these effects can be positive and negative, intended and unintended (1).

## Implementation research

Implementation research “seeks to understand and work in real-world or usual practice settings, paying particular attention to the audience that will use the research, the context in which implementation occurs, and the factors that influence implementation” (6). For the purposes of this Guide, we will define implementation research as the assessment of the uptake, institutionalization and sustainability of the evidence-based digital health intervention in a given context, including policies and practices. Implementation research optimally occurs after efficacy and effectiveness have been established, with the broader intent of informing efforts to replicate and/or expand implementation of the intervention. In practice, however, this may not occur in a linear fashion and many digital health systems may be scaled up from a prototype stage of development, bypassing traditional hurdles of efficacy and effectiveness studies.

For evaluations of digital health interventions, adoption of a hybrid approach which blends effectiveness and implementation trial elements may be warranted in cases where there is an underlying assumption of the intervention's effectiveness and/or the effectiveness of the implementation strategy, and where the risks to human subjects are minimal (see Box 4.2). Adoption of this approach may optimize the evidence collected and increase the speed at which knowledge can be translated into action (10).

#### Box 4.2. A hybrid approach to evaluation of digital health interventions

To generate evidence of effectiveness for a large-scale digital health intervention, a hybrid study design may be most appropriate; this type of study considers the effects of both the clinical intervention and the delivery/implementation processes in a real-world setting. Curran et al. 2012 outline three primary types of hybrid trial designs (10):

- **Type 1** – tests the effectiveness of an intervention on key outcomes while observing/gathering information on the context of implementation.
- **Type 2** – tests the effectiveness of both the intervention and implementation strategy on key outcomes simultaneously.
- **Type 3** – tests the effectiveness of the implementation strategy while observing/gathering information on the intervention's effect on key outcomes.



**Implementation research:** Research that “seeks to understand and work in real-world or usual practice settings, paying particular attention to the audience that will use the research, the context in which implementation occurs, and the factors that influence implementation” (6). For the purposes of this Guide, we will define implementation research as the assessment of the uptake, integration and sustainability of the evidence-based digital health intervention in a given context, including policies and practices.

**Formative evaluations:** Studies aimed at informing the development and design of effective intervention strategies. They may be conducted before or during implementation (7).

**Summative evaluations:** Studies conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced (1).

**Experimental studies:** Studies that aim to assess the effects of a treatment or intervention that has been intentionally introduced on an outcome or outcomes of interest (e.g. randomized controlled trials and quasi-experimental studies).

**Randomized controlled trial (RCT):** A type of experimental study designed to assess the efficacy or effectiveness of an intervention by comparing the results in a group of subjects receiving the intervention to the results in a control group, where allocation to the intervention and control groups has been achieved by randomization.

**Observational studies:** Non-experimental studies in which “the investigator does not intervene but rather simply ‘observes’ and assesses the strength of the relationship between an exposure and disease variable” (8).

**Hierarchy of study designs:** A ranking of study designs from highest to lowest based on their potential to eliminate bias (9).

## Formative versus summative evaluations

Once the intervention's stage of maturity has been defined (see Chapter 1, Part 1a, Figure 1.2), the programme manager needs to decide which type of evaluation is most appropriate for the evidence needs. While there are many different types of evaluations, they may broadly be classified into two categories: formative or summative. Table 4.1 provides a basic overview of types of formative and summative evaluations. This part of the chapter will focus on summative types of evaluation.



**Formative evaluations:** The two most common types of formative evaluations are needs assessments and process evaluations. Needs assessments are typically conducted before the start of an intervention to improve understanding of the needs of the intended programme clients or beneficiaries, so that the programme can be designed to best meet these needs. By comparison, process evaluations are conducted at a particular point (e.g. one year after launch) or at regular intervals during implementation to measure outputs attributed to intervention activities and inputs. Some types of formative evaluation were discussed in greater detail in Chapter 3: Monitoring (i.e. process monitoring and fidelity monitoring); however, needs assessment is not discussed in this Guide, since it is assumed that such assessment was done prior to embarking on the intervention.

**Summative evaluations:** These aim to document the broader consequences of a programme in terms of effect on key outcomes; types of summative evaluations include outcome and impact evaluations, among others (7). Outcome evaluations are concerned with the immediate and intermediate changes in key outcomes, including knowledge, awareness, coverage and behaviour change. Impact evaluations measure the long-term effectiveness of the programme in terms of effect on key health outcomes, such as mortality, morbidity and disease risk.

**Table 4.1. Formative versus summative evaluations**

Objectives		Illustrative questions asked
<b>FORMATIVE</b>		
<b>Needs assessment</b>	Determines who needs the digital health intervention, how great their need is, and what activities will best address those needs	What are the client needs? What intervention activities will best address these needs?
<b>Process evaluation<sup>a</sup></b>	Measures outputs attributed to intervention activities and inputs; this can be done continuously or as a one-time assessment	Is the intervention operating as intended?
<b>Implementation evaluation<sup>a</sup></b>	Monitors the fidelity of the intervention or technology system	Is implementation occurring in accordance with original study protocols?
<b>SUMMATIVE</b>		
<b>Performance or outcome evaluation</b>	Measures the effectiveness of intervention activities on immediate and intermediate changes in key outcomes, including knowledge, service provision, utilization and coverage	<p><b>Provision:</b> Are the services available? What is the intervention's effect on changes in service delivery?</p> <p><b>Utilization:</b> Are the services being used?</p> <p><b>Coverage:</b> Did the digital health system increase coverage of the health intervention? Is the target population being reached?</p>
<b>Impact evaluation</b>	Measures the long-term net effects or impact of the intervention on key health outcomes, including mortality, morbidity and disease risk, at the community level or higher	Were there improvements in disease or mortality patterns, or health-related behaviours?
<b>Economic evaluation</b>	Aims to determine a probable value for money from an investment	What is the incremental cost–effectiveness of the digital health intervention as compared to existing services?
<b>Secondary analysis</b>	Analysis of existing data to explore new research questions or methods not previously explored	Using the database from the International Telecommunication Union (ITU), are there associations between mobile phone ownership and women's literacy? (Questions should be tailored to research objectives)
<b>Meta-analysis</b>	Aims to integrate evidence on the effects (impact) of multiple interventions on key outcomes of interest	Overall, across multiple studies, what is the effectiveness (or impact) of this type of intervention on an outcome of interest?

<sup>a</sup> Covered in detail in Chapter 3: Monitoring (see information on process monitoring and fidelity monitoring).

Source: CDC, undated (7).

**Table 4.2. Types of study designs**

		Description	Advantages	Limitations	
Analytic	Experimental	<b>Randomized controlled trials (RCTs)</b> <ul style="list-style-type: none"><li>Individually randomized</li><li>Cluster randomized<ul style="list-style-type: none"><li>parallel</li><li>crossover</li><li>stepped-wedge</li></ul></li></ul>	<ul style="list-style-type: none"><li>A planned experiment designed to assess the efficacy of an intervention in human beings by comparing the intervention to a control condition</li><li>Allocation to intervention or control is determined purely by chance</li></ul>	<ul style="list-style-type: none"><li>Gold standard in terms of study design</li></ul>	<ul style="list-style-type: none"><li>Ethical considerations</li><li>Difficulty of randomizing subjects</li><li>Inability to randomize by locations</li><li>Small available sample size</li></ul>
		<b>Quasi-experimental studies</b> <ul style="list-style-type: none"><li>Without control groups</li><li>With control groups but no pretest</li><li>With control groups and pretests</li><li>Interrupted time-series designs</li></ul>	<ul style="list-style-type: none"><li>Aim to demonstrate causality between an intervention and an outcome but do not use randomization</li></ul>	<ul style="list-style-type: none"><li>Can be used when only a small sample size is available and randomization is not possible</li><li>Can be logistically easier to execute than an RCT</li><li>Minimizes threats to ecological validity</li><li>Can allow for population-level generalization of findings</li><li>Using self-selected groups may minimize ethical and other concerns</li></ul>	<ul style="list-style-type: none"><li>Lack of random assignment</li></ul>
	Observational	<b>Cohort</b>	<ul style="list-style-type: none"><li>Longitudinal study</li><li>Measures events in chronological order</li><li>Used to study disease incidence, causes and prognosis</li></ul>	<ul style="list-style-type: none"><li>Conducted prospectively or retrospectively</li></ul>	<ul style="list-style-type: none"><li>Can be challenging to retain individuals in the cohort over time</li><li>Lack of random assignment</li></ul>
		<b>Cross-sectional</b>	<ul style="list-style-type: none"><li>Examines the relationship between a characteristic of interest and other variables as they exist in a defined population at one single time point</li></ul>	<ul style="list-style-type: none"><li>Can be less expensive than alternatives</li></ul>	<ul style="list-style-type: none"><li>Does not establish causality</li><li>Recall bias susceptibility</li><li>Confounders may be unequally distributed</li></ul>
		<b>Case-control</b>	<ul style="list-style-type: none"><li>Retrospective studies in which two groups differing in an outcome are identified and compared based on a supposed causal attribute</li></ul>	<ul style="list-style-type: none"><li>Can be relatively inexpensive and shorter in duration than alternatives</li></ul>	<ul style="list-style-type: none"><li>Results can be confounded by other factors</li><li>Can be difficult to establish timeline of exposure</li></ul>
	Descriptive	<b>Surveillance</b>		<ul style="list-style-type: none"><li>Systematic collection, analysis and interpretation of health data</li><li>Can be active or passive</li></ul>	<ul style="list-style-type: none"><li>Provide ongoing, systematic information that is essential for planning, and service delivery</li></ul>
<b>Cross-sectional surveys</b>		<ul style="list-style-type: none"><li>Describes a health or other characteristic of interest of a population at a single time point</li></ul>	<ul style="list-style-type: none"><li>Can be less expensive than alternatives</li></ul>	<ul style="list-style-type: none"><li>Both exposure and outcome are ascertained at the same time</li><li>Do not give an indication of the sequence of events because they are carried out at one time point</li></ul>	
<b>Ecological correlational studies</b>		<ul style="list-style-type: none"><li>Look for associations between exposures and outcomes in a population rather than in individuals</li></ul>	<ul style="list-style-type: none"><li>Can be less expensive than alternatives</li></ul>	<ul style="list-style-type: none"><li>Cannot link exposure to outcome in individuals</li><li>Can be difficult to control for confounding</li></ul>	
<b>Case report</b>		<ul style="list-style-type: none"><li>Report of an event, unusual disease or association, which aims to prompt future research using more rigorous study designs</li></ul>	<ul style="list-style-type: none"><li>Can be used to spur subsequent research</li></ul>	<ul style="list-style-type: none"><li>Least publishable unit in the medical literature</li></ul>	
<b>Case-series reports</b>		<ul style="list-style-type: none"><li>Aggregates individual cases in one report</li></ul>	<ul style="list-style-type: none"><li>Can be used to spur subsequent research</li></ul>	<ul style="list-style-type: none"><li>Difficult to publish in the medical literature</li></ul>	

Source: adapted from Last, 1988 (13) and Gordis, 2014 (14).

## Study designs

Study designs aim to inform decision-making on evidence generation and the scope of monitoring and evaluation (M&E) activities. In this section, we introduce the broad classes across two types of study designs: (i) descriptive and (ii) analytic.

### i. Descriptive studies

Descriptive studies are “concerned with and designed only to describe the existing distribution of variables, without regard to causal or other hypotheses” (13). Descriptive studies aim to define the “who, what, when and where” of observed phenomena. There are two main types of descriptive studies: those concerned with individuals (case-series reports, cross-sectional studies and surveillance); and those relating to populations (ecological correlational studies) (12) (see Figure 4.1 and Table 4.2). Both types of descriptive studies may include qualitative research, study designs for which are considered in Table 4.3.

**Table 4.3. Qualitative research – study designs**

Qualitative research design	Description	Output
<b>Case study</b>	In-depth study of a case, where a case may be an individual, an event, a group or an institution	In-depth description of case
<b>Grounded theory</b>	Collected data are used to theorize about how groups work or solve problems	Theory, supported by data
<b>Phenomenology</b>	Description of lived experiences of those who have experienced the phenomenon of interest	Findings described from subject’s point of view
<b>Ethnography</b>	Close field observation (typically of a community) to describe sociocultural phenomena and characteristics	Description of culture
<b>Historical</b>	Systematic collection and objective evaluation of data from the past to inform understanding of current events/circumstances and to anticipate future effects	Biography, chronology, issue papers

Source: adapted from Donalek, 2004; Lindquist, undated; Neill, 2006 (15–17).

### ii. Analytic studies

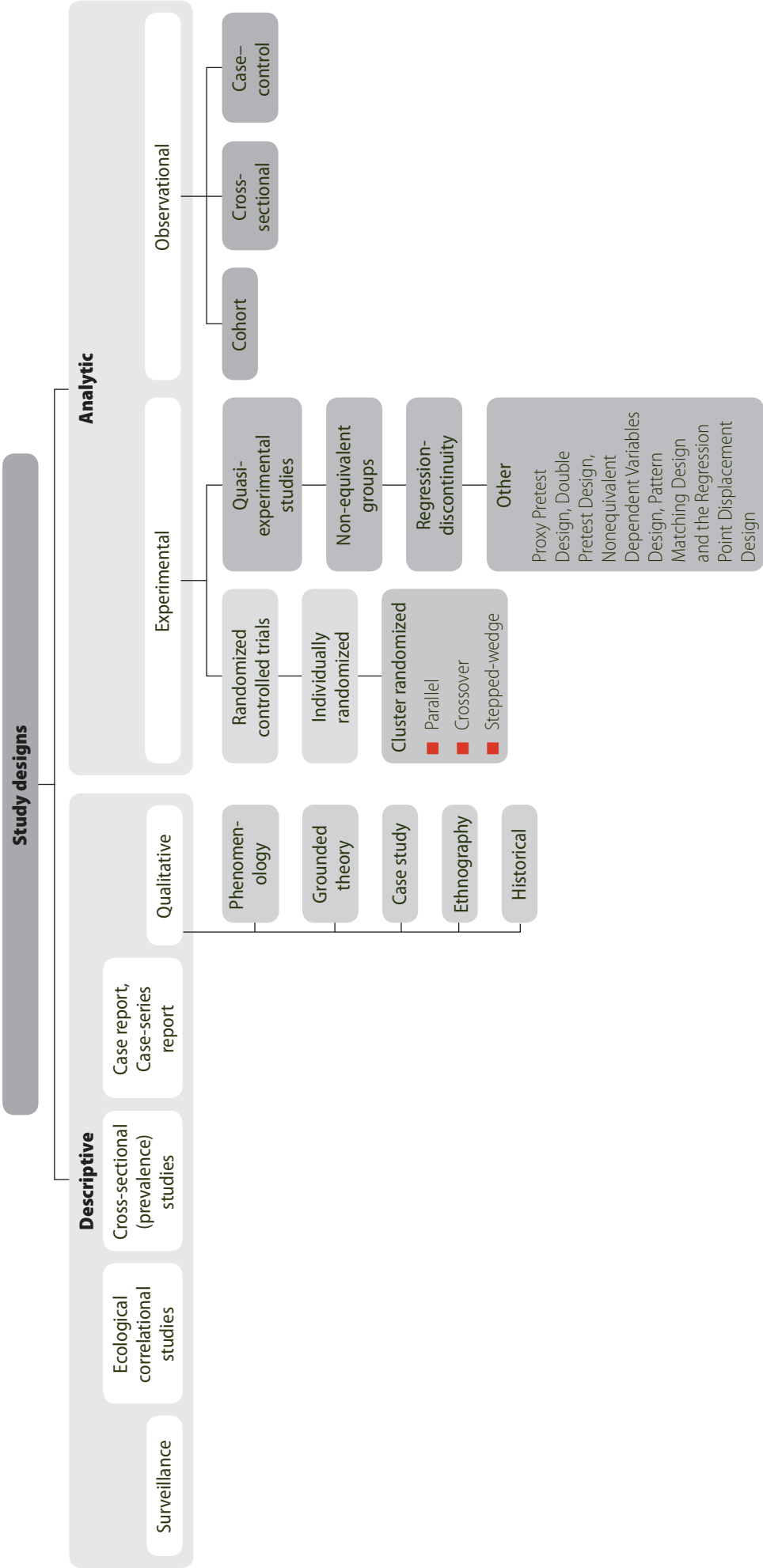
Analytic studies aim to quantify the relationship between the intervention and the outcome(s) of interest, usually with the specific aim of demonstrating a causative link between the two. These studies are designed to test hypotheses that have usually been generated from descriptive studies. There are two main categories of analytic studies: (a) experimental and (b) observational (see Table 4.2).

#### a. Experimental studies

Experimental studies aim to assess the effects of a treatment or intervention that has been intentionally introduced on an outcome or outcomes of interest. Examples of experimental studies include randomized controlled trials (RCTs) and quasi-experimental studies.

An RCT is a type of experimental study designed to assess the efficacy or effectiveness of an intervention by comparing the results in a group of subjects receiving the intervention with the results in a control group, where allocation to the intervention and control groups has been achieved by randomization. Randomization is done to avoid selection bias, improve the comparability of the groups of subjects, and largely remove the risk of any confounding effect that may be caused by unobserved or unmeasured exposures. In other words, in an RCT, the only thing that differs between the two groups is the exposure to the intervention – in this case a digital health intervention; it can be assumed that anything else that happens in the communities over the study period will likely affect both groups equally. Random assignment to the intervention or control group may be done at the level of individual participants or at the level of clusters of participants based on political boundaries (e.g. villages or hamlets). The RCT is often considered the most robust study design to demonstrate with confidence that a specific intervention has resulted in a change in a process or a health outcome.

Figure 4.1. Classes of descriptive and analytic study design

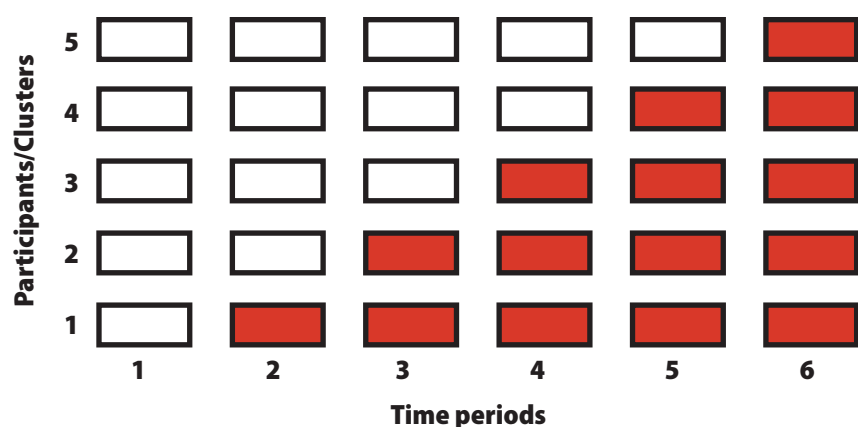


Source: adapted from Brown & Lilford, 2006; Grimes & Schulz, 2002 (11, 12).

**Table 4.4. Hierarchy of evidence by stage of evaluation**

Stage of evaluation					
		FEASIBILITY/USABILITY	EFFICACY	EFFECTIVENESS	IMPLEMENTATION SCIENCE
Confidence in the strength of evidence	Excellent	<ul style="list-style-type: none"> <li>Multicentre randomized controlled trials (RCTs)</li> </ul>	<ul style="list-style-type: none"> <li>Multicentre RCTs</li> </ul>	<ul style="list-style-type: none"> <li>Multicentre RCTs</li> </ul>	<ul style="list-style-type: none"> <li>Multicentre/quasi-experimental studies</li> </ul>
	Good	<b>1. RCT</b> <b>2. Quasi-experimental studies</b> <ul style="list-style-type: none"> <li>Interrupted time series</li> <li>With control groups and baselines</li> <li>With control groups but no baseline</li> <li>Without control groups</li> </ul> <b>3. Observational studies</b> <ul style="list-style-type: none"> <li>Cohort</li> <li>Case-control</li> </ul>	<b>1. RCT</b> <b>2. Quasi-experimental studies</b> <ul style="list-style-type: none"> <li>Interrupted time series</li> <li>With control groups and baselines</li> <li>With control groups but no baseline</li> <li>Without control groups</li> </ul> <b>3. Observational studies</b> <ul style="list-style-type: none"> <li>Cohort</li> <li>Case-control</li> </ul>	<b>1. RCT</b> <b>2. Quasi-experimental studies</b> <ul style="list-style-type: none"> <li>Interrupted time series</li> <li>With control groups and baselines</li> <li>With control groups but no baseline</li> <li>Without control groups</li> </ul> <b>3. Observational studies</b> <ul style="list-style-type: none"> <li>Cohort</li> <li>Case-control</li> </ul>	<b>1. Quasi-experimental studies</b> <ul style="list-style-type: none"> <li>Interrupted time series</li> <li>With control groups and baselines</li> <li>With control groups but no baseline</li> <li>Without control groups</li> </ul> <b>2. Observational studies</b> <ul style="list-style-type: none"> <li>Cohort</li> <li>Case-control</li> </ul>
	Fair	<b>Descriptive studies</b> <ul style="list-style-type: none"> <li>Surveillance</li> <li>Cross-sectional studies</li> <li>Ecological studies</li> <li>Case-series report</li> <li>Case studies</li> </ul>	<b>Descriptive studies</b> <ul style="list-style-type: none"> <li>Surveillance</li> <li>Cross-sectional studies</li> <li>Ecological studies</li> <li>Case-series report</li> <li>Case studies</li> </ul>	<b>Descriptive studies</b> <ul style="list-style-type: none"> <li>Surveillance</li> <li>Cross-sectional studies</li> <li>Ecological studies</li> <li>Case-series report</li> <li>Case studies</li> </ul>	<b>Descriptive studies</b> <ul style="list-style-type: none"> <li>Surveillance</li> <li>Cross-sectional studies</li> <li>Ecological studies</li> <li>Case-series report</li> <li>Case studies</li> </ul>
	Poor	<ul style="list-style-type: none"> <li>Editorials</li> <li>Expert opinion</li> </ul>	<ul style="list-style-type: none"> <li>Editorials</li> <li>Expert opinion</li> </ul>	<ul style="list-style-type: none"> <li>Editorials</li> <li>Expert opinion</li> </ul>	<ul style="list-style-type: none"> <li>Editorials</li> <li>Expert opinion</li> </ul>

**Figure 4.2. Stepped-wedge study design**



Source: Brown and Lilford, 2006 (11).

Examples of more complex RCTs include stepped-wedge, parallel or crossover study designs; each distinguished by how randomization is executed. In a stepped-wedge design, the intervention is rolled out sequentially to participants or clusters of participants over a number of time periods, with the aim that all participants will be receiving the service by the end of the study period (see Figure 4.2) (11).

When compared to a traditional parallel design, stepped-wedge study designs are considered advantageous when (a) the intervention will do more good than harm, such that limiting exposure could be unethical; and (b) there are logistical, practical or financial constraints which require the intervention to be implemented in stages (11). However, stepped-wedge designs can also have several practical challenges, including preventing “contamination” between intervention participants and those waiting for the intervention (i.e. participants who were not yet meant to be exposed to the intervention become exposed via acquaintance or contact with people already receiving the intervention), and in some instances stepped-wedge studies may require a longer overall study period than a traditional parallel design (11). Furthermore, a stepped-wedge design may not be appropriate in instances where the programme itself is likely to change over time, in response to contextual adaptations or other factors. Analysis of this kind of design is also quite complex and requires sophisticated statistical methods.

A quasi-experimental design is similar to an RCT in that it aims to demonstrate causality between an intervention and an outcome. It lacks one key feature, however: random assignment. Quasi-experimental designs are used most commonly when it is not logistically feasible or ethical to conduct an RCT (18). Because of the lack of random assignment, quasi-experimental study designs may be considered inferior, particularly with respect to internal validity. Examples of quasi-experimental studies include: interrupted time-series designs; those that use control groups and baseline assessments; those that use control groups but no baseline assessments; and those without control groups (see Table 4.2).

### **b. Observational studies**

Observational studies are non-experimental studies in which “the investigator does not intervene but rather simply ‘observes’ and assesses the strength of the relationship between an exposure and disease variable” (8). Observational studies include cohort, case-control and cross-sectional studies. Cohort studies measure events in chronological order and are used to study disease incidence, causes and prognosis (19). Case-control studies are retrospective studies in which two groups differing in an outcome are identified and compared through data analysis based on a supposed causal attribute. Cross-sectional studies aim to assess the relationship between a disease and other variable of interest at a single time point. Usually conducted in the form of a survey, cross-sectional studies are often termed “prevalence studies” because exposure and disease are determined at one time point in a population of interest.

### **Hierarchy of study designs**

Research optimally confirms and quantifies the causal relationship between an intervention and its effects (9). The level of evidence required to assess causality has traditionally been defined by the study design used (9). Hierarchies of study designs rank studies from highest to lowest based on their potential to eliminate bias (9). Traditional hierarchies of study designs focus on effectiveness studies, categorizing confidence in the strength of evidence to be *excellent* for multicentre RCTs, *good* for RCTs, *fair* for non-randomized trials, cohort studies, case-control and cross-sectional studies, and *poor* for case studies and case reports (8). Table 4.4 aims to quantify the strength of evidence for digital health interventions based on their stage of evaluation. For large- or national-scale studies that fall under the “implementation science” stage of evaluation, the conduct of more rigorous RCTs may be contraindicated or unfeasible. Instead, a quasi-experimental or observational study, inclusive of both quantitative and qualitative data collection, may be most appropriate.

## **Linking inferences with study designs and methods**

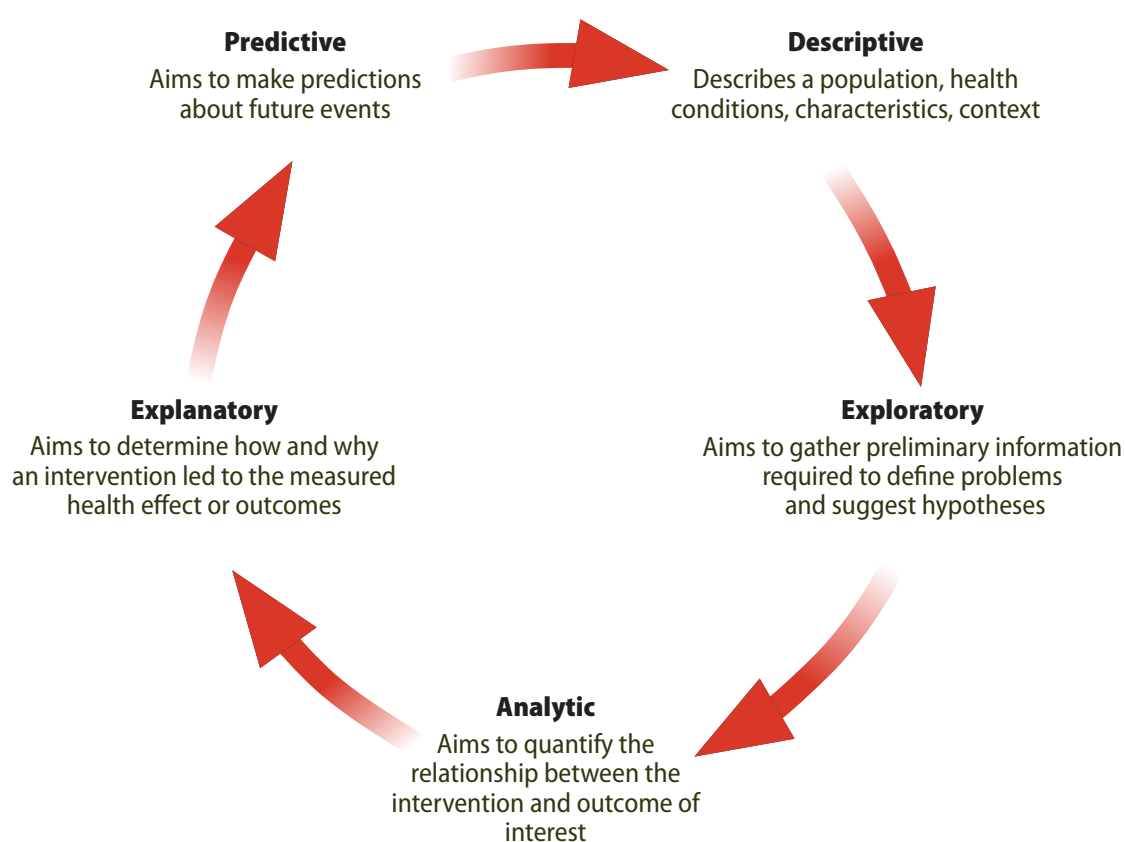
A key focus in evaluation studies is the determination of valid associations between an exposure (i.e. an intervention or treatment) and a health outcome. While there is no single best design for evaluating a digital health intervention, addressing requirements for evidence needs will involve consideration of the necessary degree of certainty. For digital health interventions that have a strong underlying evidence base that has established the intervention’s effectiveness in terms of positive health outcomes (e.g. measles immunization), an RCT may not be required to determine the effectiveness



of the immunization. However, an RCT could be undertaken to assess the comparative effectiveness of the different digital delivery strategies (e.g. SMS alerts, digitized reporting, data dashboards) in terms of improving the coverage and delivery of measles vaccinations. In the context of many large-scale programmes, the intervention of interest may account for only a portion of the variability observed in the outcomes, while socioeconomic factors, choice of delivery strategy, geographic and other contextual factors may have substantial influence. Choice of study design and accompanying research methods should be defined based on overarching research objectives and should consider how confident decision-makers need to be that the observed effects can be attributed to the intervention (20).

Some experts recommend first stratifying research questions into a limited number of categories according to strength of inference,<sup>4</sup> from descriptive and exploratory, to analytic, explanatory and predictive (Figure 4.3) (20, 21). The appropriate sequence of these may not always be linear; indeed many interventions will require predictive modelling to secure initial funding. However, categorizing research questions will help project managers to refine evidence needs, define data collection methods and better understand the limitations of the planned evaluation.

**Figure 4.3. Types of inference/study design**



Source: adapted from Habicht et al., 1999 (20) and Peters et al., 2013 (21).

Table 4.5 expands upon these five inference categories to consider research methods, study designs and the limitations of each. In the earlier section on study designs, we reviewed options for two of the most common study designs: (i) descriptive and (ii) analytic. Descriptive studies can be followed by exploratory studies, which draw upon similar data collection methodologies but aim to generate hypotheses (21). Analytic studies aim to quantify the relationship between the intervention and outcome of interest, and they include experimental and observational studies. Analytic studies explore inferences about the intervention's . . .

- adequacy (Have intervention activities met the expected objectives?)
- plausibility (Did the intervention have an effect above and beyond other external influences?) and/or
- probability (Did the intervention have an effect ( $P < x\%$ )?) (20).

<sup>4</sup> Inferences refer to the “process of determining the value or worth of something by judging it against explicit, pre-determined standards” (9).

**Table 4.5. Linking inferences with study designs and methods**

Study type/ inferences		Description	Research question(s)	Limitations	Study designs	Research methods
Descriptive		Describes a population, health conditions, characteristics, context	What are the population characteristics and needs in this context?	Does not tell you “why”; limited to a description of population characteristics	<ul style="list-style-type: none"> <li>■ Case reports/ case-series reports</li> <li>■ Surveillance</li> <li>■ Cross-sectional surveys</li> <li>■ Ecological studies</li> </ul>	<p><b>Qualitative methods:</b> Focus group discussions (FGDs), in-depth interviews (IDIs), case studies, ethnography</p> <p><b>Quantitative:</b> Surveillance, ecological correlational studies, cross-sectional surveys</p> <p><b>Mixed methods:</b> Combining qualitative and quantitative</p>
Exploratory		Aims to gather preliminary information required to define problems and suggest hypotheses	What hypotheses may explain the trends observed?	Does not assess the effect(s) of an intervention on an outcome	<ul style="list-style-type: none"> <li>■ Case reports/ case-series reports</li> <li>■ Surveillance</li> <li>■ Cross-sectional surveys</li> <li>■ Ecological studies</li> </ul>	<p><b>Qualitative methods:</b> FGDs, IDIs, case studies, ethnography</p> <p><b>Quantitative:</b> Cross-sectional surveys</p> <p><b>Mixed methods:</b> Combining qualitative and quantitative</p>
Analytic	Adequacy	<ul style="list-style-type: none"> <li>■ Compares the performance or impact of the intervention with previously established adequacy criteria (20)</li> <li>■ Assesses how well intervention activities have met expected objectives (20)</li> </ul>	Did the expected changes occur?	No control groups	<ul style="list-style-type: none"> <li>■ Observational studies</li> </ul>	<b>Quantitative:</b> Participatory action research (PAR), cross-sectional surveys, longitudinal surveys
	Plausibility	<ul style="list-style-type: none"> <li>■ Aims to determine whether the programme caused the observed effect(s)</li> <li>■ Includes a comparison group</li> <li>■ More robust than adequacy evaluations</li> </ul>	Did the programme have an effect above and beyond other external influences?	Control groups are not randomly selected, thus cannot completely rule out alternative explanations for observed differences	<ul style="list-style-type: none"> <li>■ Quasi-experimental studies</li> <li>■ Observational studies</li> </ul>	<b>Quantitative:</b> PAR, cross-sectional surveys, longitudinal surveys
	Probability	<ul style="list-style-type: none"> <li>■ Aims to determine whether the programme caused the observed effect(s) while ensuring that there is only a “small known probability that the difference between programme and control areas were due to confounding, bias, or chance” (20)</li> <li>■ Requires randomization of treatment and control activities</li> </ul>	Did the intervention have an effect ( $P < x\%$ )? (20)	Requires randomization; evaluation must start early in the intervention	<ul style="list-style-type: none"> <li>■ Randomized controlled trials (RCTs)</li> <li>■ Quasi-experimental studies</li> </ul>	<b>Quantitative:</b> PAR, longitudinal surveys, repeated cross-sectional surveys

Study type/ inferences	Description	Research question(s)	Limitations	Study designs	Research methods
Explanatory	Aims to determine how and why an intervention led to the measured health effect(s) or outcome(s)	How did the implementation of the intervention lead to the effect(s) observed?		<ul style="list-style-type: none"> <li>■ RCTs</li> <li>■ Quasi-experimental studies</li> <li>■ Descriptive</li> </ul>	<p><b>Mixed methods:</b> Quantitative and qualitative</p> <p><b>Quantitative:</b> Repeated measures of context, actors, depth and breadth of implementation across subunits; can use designs for confirmatory inferences</p> <p><b>Qualitative methods:</b> FGDs, IDIs, case studies, ethnography</p>
Predictive	Draws upon primary and secondary data to make predications about future events	What is the likely effect of programme activities on future events?	Assumption-based modelling	<ul style="list-style-type: none"> <li>■ Predictive</li> <li>■ Secondary analysis</li> </ul>	<p><b>Quantitative:</b> Agent-based modelling, simulation and forecasting modelling, data extrapolation and sensitivity analysis (trend analysis, econometric modelling)</p> <p><b>Qualitative:</b> Scenario-building exercises, Delphi techniques from opinion leaders</p>

Source: adapted from Habicht et al., 1999 (20) and Peters et al., 2013 (21).

They do not, however, answer questions about “how” or “why” the implementation of the intervention led to the observed effects; these questions are best explored by explanatory studies, ideally through a combination of quantitative and qualitative methods (21). In instances where an intervention has recently been completed such that there is no further opportunity for primary data collection, the conduct of secondary data analysis to define new hypotheses (exploratory) or make predictions about future events (predictive) may be warranted. For digital health intervention evaluations, analytic and/or explanatory inferences are likely to be the most common.

## Part 4b: Evaluation methods

In this part of the chapter, evaluation methods are grouped into four categories: (i) qualitative approaches, (ii) quantitative approaches, (iii) mixed-methods, (iv) financial and economic evaluations. In brief, qualitative approaches yield data in text, image or audio/video formats from which patterns or themes related to participants' attributes (e.g. views, behaviours) may be detected and analysed using methods and tools particular to qualitative data (22–27). Quantitative approaches (e.g. survey data collection) yield numeric data that can be analysed using statistical methods (23–30). Mixed-method approaches are a combination of both of these, involving the collection of both qualitative and quantitative data, either sequentially or concurrently, with the intent of informing the research question (24, 31). Financial evaluations assess financing and affordability, while economic evaluations aim to determine a probable value for money from an investment. The sections below provide a more detailed introduction to these methodologies, as well as comparisons among them and recommendations for further study. Part 4c presents considerations for selecting the appropriate evaluation methods for your evaluation study.

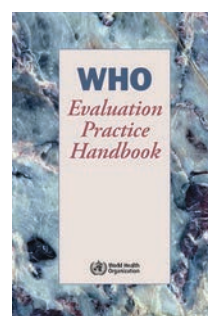


### FURTHER READING

This chapter seeks to complement other existing notable resources on conducting evaluations, including the following:

WHO evaluation practice handbook. Geneva: World Health Organization; 2013 ([http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf)).

Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. Impact evaluation in practice. Washington (DC): The World Bank; 2011 ([http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact\\_Evaluation\\_in\\_Practice.pdf](http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf)).



### KEY TERMS

**Focus group discussions (FGDs):** A type of qualitative research method used when researchers are interested in the opinions or perspectives of a group of individuals (22–27, 32). FGDs may be used to get feedback before, during or after a project, to reach groups that are underrepresented in surveys, to compare and contrast norms between groups, and for goal-setting and prioritization (22–27, 32, 33).

**In-depth interviews (IDIs):** The process of eliciting detailed perspectives, opinions, experiences and feelings of individuals (22, 26, 27, 33). IDIs may be conducted over the phone or in person.

**Structure observations:** The process and techniques used for observing and documenting the daily experiences, actions and situations of the population of interest in their everyday environments (26).

**Participatory action research (PAR):** The use of quantitative, qualitative or mixed-methods approaches in a manner that prioritizes the role of participants in all aspects of research and implementation (21, 34, 35).

## Qualitative methods

Qualitative research methods are suitable for instances where diverse and in-depth understanding of participants' knowledge, behaviours, opinions, attitudes, values, motivations and sociocultural contexts is desired. Qualitative research methods can be useful not only to inform the initial design of digital health interventions, but also to evaluate these interventions. The purpose of qualitative research is exploration (22, 24, 26). The five main types of qualitative research designs were summarized in Table 4.3 earlier in this chapter. Qualitative methods are often used to generate hypotheses,

explore certain research findings further, or provide context for quantitative data (22, 24–27). Data obtained from qualitative approaches are rich in nature and can be in text, audio, image or video format. Objects and documents relevant to the research question may also be collected during qualitative research (26). Unlike quantitative methods, which use structured questionnaires with closed-ended (i.e. yes/no or multiple-choice) questions (e.g. “Did you seek antenatal care during your last pregnancy?”), qualitative methods employ guides with open-ended questions and prompts (e.g. “Tell me about your experience seeking antenatal care during your last pregnancy”) (23–26), allowing participants to construct responses in their own words.

The three qualitative methods most commonly used in the evaluation of digital health interventions are (i) observation, (ii) in-depth interviews (IDIs) and (iii) focus group discussions (FGDs).

### **i. Structure observations**

Observation refers to the process and techniques used for observing and documenting the daily experiences, actions and situations of the population of interest in their everyday environments (26). Observation is useful when details on processes, protocols or workflows are desired to guide the development of a digital health intervention or to measure changes in process efficiencies as part of evaluation efforts. Observation is also useful for capturing nuances of human interactions (e.g. patient–provider or provider–provider interactions). Examples of observation may include:

- Attending and observing the proceedings of a vaccination outreach clinic for a day to understand supply- and demand-side facilitators and barriers to childhood vaccination: The data may be used to determine the features of a digital health system (e.g. notifying mothers using a text message when the vaccinator is on site) being designed to facilitate routine vaccinations.
- Shadowing a community health worker (CHW) making home visits to observe their interactions with clients to understand processes and workflows: The data may be used to create digital job aids for CHWs to facilitate their interactions with clients (e.g. to assist in providing health education).

During observation, researchers take thorough, objective notes and may interact informally with the people being observed to ask clarifying questions related to their actions, interactions or behaviours (26). The steps involved in conducting observation include:

#### **Before:**

- a. Identify (based on the research question to be addressed):
  - ✓ the population to be observed
  - ✓ the activities, interactions or behaviours to be observed.
- b. Determine the venue(s) where observation is to be conducted.
- c. Secure permission from relevant entities for conducting observation, including any special permissions required for taking photos, capturing audio/video recordings and collection of materials.
- d. Select the day and set the duration for the observation.

#### **During:**

- e. Arrive at the site and take detailed, objective notes. Pose clarifying questions as needed, but without excessively interrupting the proceedings. It is good practice to make a note of follow-up questions that arise so those may be explored later on.
- f. Collect any relevant audio and/or video recordings, images, documents or objects.

#### **After:**

- g. Expand shorthand notes to create narrative field notes.
- h. Analyse the data to draw out themes and concepts relevant to the research question (26).

Structured observations are unique in that the researcher is immersed in the environment and activities of the population of interest instead of bringing individuals from a study sample to a research office or base. As such, observation has the following advantages:

- It allows researchers to gather a breadth of information beyond what may be available through the responses of study participants during in-depth interviews (IDIs) and focus group discussions (FGDs). During IDIs and FGDs, respondents may omit details that they think are irrelevant to the research and details on aspects of their work or life that are so routine or ingrained that they rarely pay attention to them anymore.
- It allows researchers to compare the subjective reports of members of the target population (i.e. what they say happens) with reality (i.e. what is observed to happen). For example, a health supervisor may indicate that vaccination clinics are open daily from 9 am to 5 pm, but the researcher may find that the clinics are open only between 10 am and 12 noon on most days, due to long travel times and competing job responsibilities among the vaccinators.
- It informs future research and the design of questions to be used for other study methods, such as FGDs. It is often useful to conduct participant observation prior to other modes of inquiry so that researchers can follow up on any questions that arise during observation (26).

The disadvantages of observations include:

- The method is time- and resource-intensive, requiring the researchers to travel to the site where members of the target population are located and spend days or weeks with them. In ethnographic research, it is not uncommon to spend a year or more observing populations of interest (26).
- Documentation of proceedings requires extra effort. The researchers must take detailed notes and/or arrange audio/visual recordings at the time of observation and rewrite notes and/or transcribe/summarize details from recordings. While doing so, they must make an effort to remain objective, as subjectivity may affect the interpretation of data during analysis (see Box 4.3 for an example of objective versus subjective note-taking) (26).

#### **Box 4.3. Subjective versus objective note-taking during observation**

**Subjective observation:** A few young children were at the vaccination clinic.

**Objective observation:** Four children between the ages of 2–4 years were seated in the vaccination clinic's waiting area.

## **ii. In-depth interviews**

In-depth interviews (IDIs) refer to the process of eliciting detailed perspectives, opinions, experiences and feelings of individuals (as opposed to FGDs, which involve groups of individuals) (22, 26, 27, 33). IDIs may be conducted over the phone or in person. IDIs may be used to:

- Interview experts on the research topic of interest. For instance, an IDI with a ministry of health representative on policies related to antenatal care (ANC) programmes in the region may help determine common challenges and priority needs, as well as ANC interventions that are included in a digital health intervention targeting pregnant women.
- Interview individuals on sensitive topics which they may be reluctant to discuss in group settings. For instance, IDIs with a series of pregnant women with HIV/AIDS may inform the features of a digital health intervention to prevent mother-to-child transmission (PMTCT).

IDIs are often used to inform the development of user personas that are used to understand needs and refine the expected use of a digital intervention. The process involves questioning, prompting, listening and evaluating (32). Typically, interview questions and prompts are listed in an interview guide. IDIs can last anywhere from 30 minutes to two hours and are audio-recorded to assist with documentation. In addition, the interviewer takes shorthand notes to supplement the audio-recorded information with contextual information or observations (26). Notes are also used to remind the interviewer of issues that may need further clarification or which may inform future research (26). Finally, the notes may serve as a backup if the audio-recording fails unexpectedly. Following the interview, the shorthand notes are expanded and the audio-recording is transcribed for analysis (26).



It should be noted that the goal of the IDI is not to arrive at one single answer to the research question, but rather to obtain an exhaustive view of the individual's opinions on the topic (22, 23, 26, 27, 33). The advantage of an IDI is that it allows the discussion of topics in a private setting. The disadvantage is that it only provides the viewpoint of one individual at a time. Due to the small number of individuals interviewed, the data, although informative, are not generalizable (22–27, 32).

### iii. Focus group discussions

Focus group discussions (FGDs) are used when researchers are interested in the opinions or perspectives of a group of individuals (e.g. pregnant women, adolescent girls) (22–27, 32). FGDs may be used to get feedback before, during or after a project, to reach groups that are underrepresented in surveys, to compare and contrast norms between groups, and for goal-setting and prioritization (22–27, 32, 33). As with IDIs, the purpose of FGDs is not to arrive at a consensus but rather to gather a diversity of views that can inform the research question. During FGDs, participants are influenced by group dynamics. It is typical for FGD participants to be selected based on some shared characteristic (e.g. they are in the same age range, of the same gender, or have the same health concerns) (22–27, 32, 33). Examples may include:

- FGDs with pregnant women to understand cultural norms surrounding birth.
- FGDs with new mothers on their attitudes towards vaccinations.
- FGDs with adolescent girls on contraceptive use.

FGDs typically have two or more facilitators, including a moderator and a note-taker (assistants may also include translators, additional note-takers) (26). During an FGD, the moderator typically leads a group of approximately 6–12 individuals through a discussion to elicit their views on a particular topic. Moderators follow a guide, and use questioning, prompting, listening and evaluating to encourage responses from participants. Questions start out broadly and then get more specific as participants open up (see Box 4.4 for example questions) (26). Participants are encouraged to expand upon their responses so that in-depth information may be captured. FGDs typically last 60–90 minutes and no more than two hours (26). The discussion is audio-taped and later transcribed. The note-taker may include notes about participants' non-verbal cues (e.g. body language, emotion during response), along with notes about the content of the discussion.

#### Box 4.4. Example FGD questions for community health workers using a digital health system

##### Opening question:

- What are the common challenges that you face in providing care to clients in the community?

##### Substantive questions:

- Which features of the digital health system do you like? Which ones do you dislike? What would you change about them?
- What additional features would you like to see in the digital health system?
- How would the digital health system change your ability to provide health services to clients?

##### Closing questions:

- What else would you like to tell me about your experience with the digital health system?
- Is there anything I might have missed?

### Sampling for qualitative methods

Qualitative research methods usually involve a non-representative sample of the population, hence the findings are not generalizable (26). The two most common sampling strategies employed to recruit participants for qualitative research are purposive and snowball sampling (26).

### **Purposive sampling**

With this method, researchers specify the characteristics of the individuals they would like to recruit (e.g. pregnant women). The sample size may be unspecified (this is typical for purposive sampling) or restricted to a certain number (also referred to as “quota sampling”).

### **Snowball sampling**

This method builds upon purposive sampling methods and asks participants who have already been identified or recruited to refer other individuals who could potentially participate in the study. The referred individuals may belong to the participant’s immediate family or extended social networks. Snowball sampling is usually continued until saturation of information is achieved – that is, no new information is learnt from additional interviews or FGDs with the recruited participants.

## **Recording and analysing qualitative data**

As mentioned, it is common to audio-record the proceedings of qualitative research as well as to take shorthand notes to supplement any recordings (26, 36). Once data collection is completed, the shorthand notes are expanded into field notes – a more elaborate, narrative version of the shorthand notes. It is best to expand upon notes sooner rather than later (typically this should be done on the same day as the FGD or IDI) to minimize any losses in recall. Audio-recordings are also transcribed, and the recordings and notes may also be translated into English if needed.

Once the field notes and transcripts are ready, it is common to use some form of computer-assisted qualitative data analysis software (CAQDAS) to organize the data for analysis (26, 36). Examples of CAQDAS that are commonly used include Atlas.ti, QSR-Nvivo, MAXQDA and HyperRESEARCH. While the software can help with organizing and categorizing the data into concepts or themes, the onus of drawing interpretations and extracting connections between the different concepts is on the researcher (36).

There are two primary approaches to qualitative data analysis: deductive and inductive (36).

**The deductive approach:** The research question is used to group the data and look for similarities and differences.

**The inductive approach:** An emergent framework is used to group data and examine relationships in the data.

More information on qualitative data analysis can be found in the box on evaluation resources at the end of this chapter.

## **Quantitative methods**

Quantitative research methods involve the collection and analysis of numerical data. The primary purpose of conducting evaluations using quantitative research is to establish causality or association between independent variables (e.g. age) and a dependent variable (e.g. ability to use the digital health application) (24, 28–30). Study designs that may use quantitative methods include experimental (e.g. RCTs), quasi-experimental (e.g. pretest/post-test comparisons), observational (e.g. prospective cohort) or descriptive (e.g. cross-sectional) studies (28–30), all of which are designed to arrive at a consensus with regard to the research question. Quantitative methods use structured questionnaires (e.g. polls and surveys) with closed-ended questions (e.g. “Did you use the digital health application to estimate the client’s gestational age?”) (28–30, 37). Quantitative methods are not useful when in-depth information or insight is required on the “how” or “why” of a particular issue, since participants can only choose from pre-specified responses (i.e. yes/no or multiple choice). Since data are collected from large samples, the results of quantitative research are intended to be representative of the population of interest and can be generalized beyond the sample to that population as a whole (22–25, 28). The sections below provide an introduction to polls and survey-based quantitative data collection, followed by sections with further details on sampling and analysis.

### **Polls**

Polls typically consist of a single, closed-ended question asking about the participant’s behaviour or opinion, and offering a limited choice of responses (e.g. “Did you use a bednet last night?”; response options: yes or no). Due to its simple nature, a poll offers a quick way to gather responses from a large number of respondents. Polls are not useful when comprehensive information on a given topic is desired. Polls may be implemented via text messages, interactive voice response, Unstructured Supplementary Service Data (USSD), websites, etc. Results of polls are available quickly and do not require

complex statistical analyses. Polls are useful when a quick snapshot of public opinion or behaviour is needed to inform ongoing or planned work. Polls may be used to collect data from recipients related to digital health interventions (e.g. “After receiving the text message reminder to use a bednet, did you actually use a bednet that night?”).

## Surveys

Surveys are data collection methods that utilize structured questionnaires comprising many questions to elicit comprehensive information from a sample of the population of interest (28–30, 37). In comparison to polls, surveys are more time- and resource-intensive. Surveys can include questions in multiple formats such as multiple choice, yes/no, true/false, text input. Surveys are most often used to collect quantitative data, although some qualitative data might also be obtained through the incorporation of open-ended questions.

The following steps are involved when planning and conducting a survey:

1. Determine the survey objectives.
2. Determine the target population, sample size and sampling frame.
3. Determine the mode of data collection.
4. Construct the survey (either develop new questions or adapt ones from existing surveys that address the objectives determined in step 1).
5. Pilot-test and modify the survey as needed until it is ready.
6. Recruit respondents and administer the survey.
7. Analyse the data.

Managers of digital health interventions may be interested in conducting surveys with two groups of individuals: community-based clients and health-care providers (see Table 4.6).

**Table 4.6. Reasons for conducting surveys with community-based clients and health-care providers**

Surveys of community-based clients of a digital health intervention (e.g. women, men, adolescents, children) may be conducted to . . .	Surveys of health-care providers (e.g. doctors, nurses, community health workers and their supervisors) may be conducted to . . .
■ collect data on their sociodemographic attributes or medical histories (e.g. mean age, history of tobacco use, vaccinations received)	■ collect data on their background (e.g. demographics, job title, medical education and training)
■ assess their knowledge, attitudes and practices related to health issues of interest	■ assess their knowledge, attitudes and practices related to health issues of interest
■ gather data on mobile phone ownership (e.g. proportion who own a smartphone) and usage (e.g. texting proficiency)	■ gather data on mobile phone ownership (e.g. proportion who own a smartphone) and usage (e.g. proficiency in using smartphone touchscreens)
■ understand the requirements for the digital health intervention (e.g. which features of the digital health intervention might be most useful to increase access to health information and services)	■ understand the requirements for the digital health intervention (e.g. which features of the digital health intervention might be most useful to improve service delivery)
■ assess the usability of the digital health system	■ assess the usability of the digital health system
■ assess satisfaction with the digital health intervention	■ assess satisfaction with the digital health intervention
■ evaluate project outcomes related to client health (e.g. whether certain health services such as ANC were accessed more as a result of text message reminders)	■ evaluate project outcomes related to provider performance (e.g. how the use of a mobile-phone-based decision support tool may affect adherence to clinical protocols)
	■ collect data on the health-care environment (e.g. patient populations served, clinic characteristics, menu of clinical services provided, availability of clinic infrastructure such as equipment, supplies, exam rooms)

Surveys are classified based on the mode of data collection: in-person interviews, mailed surveys, telephone surveys, online/web-based surveys (e.g. if the project has a dedicated web portal) and other methods of remote data collection such as SMS, USSD or voice-based queries (the latter are useful for surveys with few questions) (28–30, 37). The most common method of conducting surveys is by administering a questionnaire in person. Surveys administered in person may be done on paper or on a digital device, such as a tablet computer. Box 4.5 lists examples of digital data collection platforms that can be deployed on mobile phones and/or tablets. Many of these platforms offer features such as offline data storage for situations when poor connectivity may prevent data from uploading. The platforms also offer back-end infrastructure to enable visualization of collected data (e.g. tables, charts, geomaps) and the ability to export data in formats compatible with popular statistical analysis programs.

#### **Box 4.5. Examples of digital data collection platforms that can be deployed on mobile devices**

- CommCare <https://www.comcarehq.org/home/>
- Magpi <http://home.magpi.com/>
- Medic Mobile <http://medicmobile.org/>
- ONA <https://ona.io/home/>
- OpenDataKit (ODK) <https://opendatakit.org/>
- Rapid Pro <https://community.rapidpro.io/> (useful for polls)
- TextIT <https://textit.in/> (useful for polls).

### **Sample size considerations**

Surveys are conducted on a subset of the population (referred to as the sample) because it is usually logistically and financially unfeasible to survey each and every member of the population of interest (28, 38, 39). By careful selection of the sample (see the next sub-section: Sampling for survey data collection), one can ensure that the sample is representative of the population and, hence, that the survey results will be generalizable to the population. Sample size calculations help to determine how many individuals from the population must be surveyed to achieve scientific validity of results. See Box 4.6 for some factors influencing sample size calculations.

#### **Box 4.6. Factors influencing sample size**

- Statistical power of the study
- Acceptable level of statistical significance
- Anticipated effect size of primary outcome
- Event rate of primary outcome in the population
- Standard deviation of the primary outcome in the population
- Study design

### **Sampling for survey data collection**

There are two primary approaches for sampling survey respondents in a target population: probability sampling and non-probability sampling (28).

**Probability sampling:** This is a sampling method that involves a form of random selection and gives all individuals in the population an equal chance (probability) of being selected for the survey (40).

Probability sampling approaches can be further classified as follows:

- Simple random sample: Choose individuals from the population at random.
- Systematic random sample: Begin with a random individual, and choose every *i*th person you encounter.
- Stratified random sampling: Stratify (i.e. divide) your population using a particular characteristic (e.g. urban/rural), then sample randomly within the strata.
- Cluster random sampling: Divide your population based on geographic location into clusters (e.g. clusters of villages), then randomly select a fixed number of clusters and survey all respondents within the selected clusters.
- Multi-stage sampling: Use a combination of the above sampling approaches. For example, in a two-stage sampling process, we may begin with cluster random sampling then use systematic random sampling within the selected clusters.

**Non-probability sampling:** This is a sampling method that does not rely on randomization and allows higher probabilities for certain individuals to be selected for the survey. As a result, non-probability sampling techniques result in samples that are not representative of the population. This approach is generally discouraged when the purpose of the evaluation is to generalize knowledge, so it is typically more commonly used in qualitative research.

Non-probability approaches include:

- Convenience sampling: Sample the individuals who are most conveniently available or available first.
- Purposive sampling: Sample the individuals who have certain characteristics (e.g. health-care providers).
- Snowball sampling: Begin with purposive sampling and ask the sampled individuals to refer others who may be similar to them for the survey.

## Quantitative data analysis

Quantitative data analysis involves the use of statistical methods to describe the data (in tables, charts, figures, etc.), to demonstrate significant (or non-significant) differences between groups, to quantify correlations or associations between variables, and to control for confounding factors (28, 41). Tables are used to document aggregate and disaggregated data, supplemented with information on sample sizes, value of test statistics, confidence intervals and significance levels (*P* values). Advanced statistical approaches (e.g. regression) may be used to create predictive models for the causal pathway or associations between the dependent and independent variables of interest. Commonly used data analysis software includes Microsoft Excel, STATA, SAS and SPSS. These programs vary in their ability to perform complex analyses and may require training and knowledge of specific programming languages and syntaxes for use. Additional information on quantitative data analysis may be found in the box on programme evaluation resources at the end of this chapter.

Table 4.7 illustrates the differences between qualitative and quantitative approaches.

**Table 4.7. A comparison of qualitative and quantitative approaches<sup>a</sup>**

	Qualitative	Quantitative
<b>Purpose</b>	<ul style="list-style-type: none"> <li>■ Exploratory</li> <li>■ Generate hypotheses</li> <li>■ Understand social interactions, behaviours, opinions, etc.</li> </ul>	<ul style="list-style-type: none"> <li>■ Predictive</li> <li>■ Test hypotheses</li> <li>■ Examine associations and causality</li> </ul>
<b>Types of questions</b>	<ul style="list-style-type: none"> <li>■ Open-ended questions</li> </ul>	<ul style="list-style-type: none"> <li>■ Closed-ended questions</li> </ul>
<b>Types of data collection instruments</b>	<ul style="list-style-type: none"> <li>■ Observation and note-taking, recording (image/audio/video), gathering related materials</li> <li>■ Question guides allowing for flexibility in responses by participants</li> </ul>	<ul style="list-style-type: none"> <li>■ Structured data collection instrument/questionnaire with pre-determined responses</li> </ul>
<b>Examples of methods</b>	<ul style="list-style-type: none"> <li>■ Participant observation</li> <li>■ In-depth interviews (IDIs)</li> <li>■ Focus group discussions (FGDs)</li> </ul>	<ul style="list-style-type: none"> <li>■ Surveys</li> </ul>

	Qualitative	Quantitative
<b>Types of data generated</b>	■ Rich, contextual and detailed information in text, audio, image and video formats	■ Numeric data on frequency, severity, change over time and associations between variables
<b>Purpose of data analysis</b>	■ Identify patterns, features and themes in data	■ Identify levels and trends and statistical relationships in data
<b>Sample size and representation</b>	■ Sample size is small and non-representative of population	■ Sample size is large and representative of population
<b>Sampling techniques</b>	■ Non-probability sampling approaches more common	■ Probability sampling approaches more common

<sup>a</sup> Mixed-methods approaches include a combination of both.

Source: adapted from The Association of Faculties of Medicine of Canada, undated; Creswell, 2003; Xavier University Library, 2012; Mack et al., 2005; Gilson et al., 2011 (23–26, 42).

## Mixed-methods approaches

Approaches that combine the use of qualitative and quantitative modes of inquiry are termed mixed-methods approaches (24, 31). Mixed-methods evaluations involve the use of both closed-ended questions and open-ended questions or observations. The resulting data are multifaceted and may include text, numeric data, images, audio recordings and/or videos.

Examples of ways in which evaluation methods may be combined in the context of digital health interventions include:

- use of IDIs to assess the cultural acceptability of a digital health intervention in following-up findings from observation.
- use of FGDs to elaborate on reasons for low engagement with the digital health system based on findings from quantitative surveys.
- use of a theoretical framework to inform the study design and procedures; this could involve data collection in a sequential or concurrent manner depending on the framework and research objectives (24, 31).

The use of a mixed-methods approach offers several advantages (24, 31). It can help to offset the limitations of any one particular method (e.g. lack of generalizability of qualitative data and lack of breadth and depth of information in quantitative data). Since this approach involves data being collected using different methods for the same research question, it allows triangulation of results and improves the validity of the findings (42). Mixed-methods approaches are also desirable when results from one method are needed to inform research questions being developed for use by another method. Mixed-methods approaches allow the development of theories (through qualitative approaches) as well as the ability to test them (using qualitative or quantitative approaches).

Some disadvantages of using mixed methods include the complexity of research design and the time and resources needed to implement data collection and to analyse and interpret both qualitative and quantitative data in the same study.

Some considerations for the reporting of mixed-methods studies are presented in Box 4.7. Further reading on mixed methods may be found in the box on programme evaluation resources at the end of this chapter.

### Box 4.7. Checklist for the reporting of mixed-methods studies

- Justification for mixed-methods approach
- Description of design (purpose, priority and sequence of methods)
- Methods (sampling, data collection and analysis plan for each method)
- Integration of methods (where and how)
- Limitations
- Lessons learnt by mixing methods

Source: Peters et al., 2013 (21).



## Economic and financial evaluations

Financial evaluations deal with the questions of whether the organization and digital health users can afford the digital health system, and how it will be financed. Economic evaluations aim to determine a probable value for money from an investment. Your economic evaluation identifies what you should do; your financial evaluation shows if you can pay for it, and how you will pay for it.

### Financial evaluations

Financial evaluation has two goals. The first is to determine the affordability of your preferred digital health intervention by comparing net costs with cash flows over time (43), which is known as budget impact analysis (44, 45). The second goal is to develop sustainable financing plans to ensure that your project has ongoing resources available to pay for it.

Viable financing and affordability are essential for delivering and sustaining the value for money identified by your economic evaluation, so the financial evaluation should follow on from your economic evaluation. The United Kingdom's Green Book methodology provides guidance on both economic and financial evaluations (46).

### Economic evaluations

Economic evaluations allow you to determine whether your project is good value for money, by comparing two or more alternatives in terms of their inputs and their effects on a common outcome or impact.

Economic evaluations are generally undertaken using one of five main methodologies:

- Cost–effectiveness analysis (CEA)
- Cost–utility analysis (CUA)
- Cost–benefit analysis (CBA)
- Cost–consequence analysis (CCA)
- Cost-minimization analysis (CMA).

Table 4.8 outlines the characteristics of these five different methodologies. In this Guide, we focus on the three most common forms of economic evaluations, CEA, CUA and CBA, which are listed first in the table. These are described here in greater detail.

**Cost–effectiveness analysis (CEA)** estimates the difference between the costs of your resources and the effectiveness of their use. This is the most commonly used form of evaluation in health economics. Effectiveness is most frequently estimated by a single measure, such as the number of lives saved, or episodes of illness averted (47).

**Cost–utility analysis (CUA)** considers people's quality of life and the length of life they will gain as a result of an intervention. It aims to overcome the one-dimensional limitations of consequence in a CEA by using utility-based outcome units to compare different interventions (48). Quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs) are two types of summary measures that can be used to estimate cardinal values assigned to health states (49). They are both based on a scale of 0 to 1, but have inverted interpretations: whereas a score of "1" represents "perfect health" for QALYs, it represents "death" for DALYs. Use of these utility measures allows you to capture more comprehensively the full state of health and well-being. Moreover, these measures allow you to compare your digital health intervention not just to the comparator of your intervention (e.g. the status quo/usual care), but to other uses of your resources across the health sector, such as a comparison between your digital health SMS programme for pregnancy care and a tuberculosis control programme, or an HIV programme. You may find that for your initial economic evaluations, you won't have access to the health data needed to calculate QALYs or DALYs. You'll need medical advice if you want to use QALYs and DALYs in your digital health evaluation.

**Cost–benefit analysis (CBA)** measures estimated costs and benefits over time in monetary values for all types of stakeholders involved in your intervention. These values come from the prices and costs of components that have market values, and from estimated monetary values for intangible components that do not have market values, using willingness-to-pay (WTP) techniques. Because of this assignment of monetary value to consequences, CBA is arguably the most controversial

form of economic evaluation. Where benefits are measured in terms of lives saved, assigning a monetary value to these lives can elicit strong reactions. Many digital health programmes don't directly save lives – they aim to improve health care – and therefore life valuations are seldom relevant for these. Many analysts consider CBAs to be “the methodological foundation for turning theory into a pragmatic evaluation tool” (50, 51). The eHealth Impact (eHI) methodology is a version of CBA applied to digital health evaluations (52). CBA is relevant when programme managers and the broader group of decision-makers want to compare intervention results to other projects that are competing for your organization's scarce resources.

**Table 4.8. Types of economic evaluation for digital health**

	Definition	Costs	Consequences	Illustrative questions
Cost–effectiveness analysis (CEA)	Comparison of two alternatives where consequences of the intervention are measured in natural units	Monetary units	Effectiveness measure: Natural units (e.g. life-years gained, lives saved, cases detected)	Is sending SMS reminders to encourage pregnant women to come for antenatal care more cost-effective than a strategy of not sending messages?
Cost–utility analysis (CUA)	Form of CEA where consequence is measured in terms of healthy years	Monetary units	Utility measure: Quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs)	Health-care providers are sent SMS reminders to notify them if a home birth has occurred. Is this SMS strategy cost-effective compared with a strategy of not sending messages, in terms of increasing the number of healthy years of life for newborns?
Cost–benefit analysis (CBA)	Comparison of two alternatives where consequences of the intervention are measured in monetary terms	Monetary units	Benefit measure: Monetary units	Which digital health option – using SMS or interactive voice response – offers the best net benefit for each client? How does this change over the timescale of the digital health intervention?
Cost–consequence analysis (CCA)	Examination of costs and consequences but without isolating a single consequence or aggregate consequences into a single measure	Monetary units	Consequence measure: Natural units	How does the digital health intervention make a difference to the rate of: provision of perinatal care in the home to pregnant women; trained attendants present at birth; newborns fully vaccinated; adequate postnatal care for mothers; and improved education and development due to better health?
Cost-minimization analysis (CMA)	Comparison of relative costs of interventions with equivalent effects	Monetary units	Consequences are assumed to be equivalent	You want to provide tetanus toxoid immunizations to 1000 pregnant women. Which strategy will cost the least to achieve this result – (i) having a community health worker promote and refer pregnant women, or (ii) using a digital health strategy of sending SMS reminders to pregnant women to attend ANC?

## Part 4c: Which evaluation activities are right for you?

The state of maturity of the digital intervention will guide which types of evaluation are most appropriate. As described in Chapter 1, Part 1b (see Table 1.1), the stages of maturity span across the continuum from pre-prototype, prototype, pilot and demonstration, to scale-up and finally integrated and sustained programme (Table 4.9). For solutions that fall in the pre-prototype and prototype (early) stages of development, initiating an evaluation may help to determine your digital health programme's trajectory and avoid higher-risk directions. Once the basic system has been refined and stabilized, evaluation activities will be required to advance the testing of the system and the intervention and thus enable its movement across stages of evaluation.

For digital health interventions in the integration (advanced) stages of development, predictive modelling activities may be indicated to facilitate the estimation of intervention effectiveness and/or resource requirements for delivery at scale. These may either be executed from the vantage point of the intervention being considered for integration into a national health system, or more broadly as system-wide assessments through application of modelling tools, such as OneHealth (<http://www.avenirhealth.org/software-onehealth> [53]).

**Table 4.9. Linking stages of maturity with evaluation methods and claims**

Stage of maturity		Stage of evaluation
Early	<b>Pre-prototype:</b> This stage includes hypothesis building, needs/context assessment, and testing of usability/feasibility and technical stability.	<b>Feasibility:</b> Assess whether the digital health system works as intended in a given context.
	<b>Prototype:</b> During this phase, user-focused designs are created and tested, and functionality, technical stability and usability are tested in an iterative process. Ways to improve the system are examined to enhance relevance.	<b>Usability:</b> Assess whether the digital health system can be used as intended by users.
	<b>Pilot:</b> This stage examines whether the digital health intervention can produce the desired effect under controlled circumstances. The pilot project is usually a single deployment.	<b>Efficacy:</b> Assess whether the digital health intervention can achieve the intended results in a research (controlled) setting.
Mid	<b>Demonstration:</b> In this stage, the intervention is no longer taking place in controlled conditions but is still limited in terms of population/geography (usually restricted to a particular region or subregion).	<b>Effectiveness:</b> Assess whether the digital health intervention can achieve the intended results in a non-research (uncontrolled) setting.
Advanced	<b>Scale-up:</b> In this stage, approaches are ready to be optimized and scaled up across multiple subnational, national or population levels.	<b>Implementation science:</b> Assess the uptake, integration and sustainability of evidence-based digital health interventions for a given context, including policies and practices.
	<b>Integrated and sustained programme:</b> Efforts at this stage are focused on determining the necessary components of an enabling environment that will support impact of the intervention at a large scale (i.e. policies, financing, and human resources, interoperability, etc.). The intervention has been integrated into a broader health system.	

## Which evaluation activities best address your evidence needs?

### Is your evaluation formative or summative, or both?

Formative evaluations are typically conducted internally by the organization to allow for findings to influence programme design and implementation. However, for some activities, namely process evaluations, it is feasible that an external entity could execute a formative evaluation. Summative evaluations, while optimally conducted by an external entity, should be used to provide the evidence necessary to continuously inform improvements, implementation and expansion of the intervention.

### Define your study inferences

Inferences refer to the “process of determining the value or worth of something by judging it against explicit, pre-determined standards” (9). In the earlier section in Part 4a, on Linking inferences with study designs and methods, we

reviewed five main categories of inferences: descriptive, exploratory, analytic, explanatory and predictive. To define the overarching research objectives and evidence needs, you need first to identify, categorize, and then prioritize your research inferences across these categories.

Most evaluations of digital health interventions will likely focus on analytic and explanatory inferences. Among the former, we distinguished between three sub-categories of inferences: adequacy (Have programme activities met expected objectives?), plausibility (Did the programme have an effect above and beyond other external influences?), and probability (Did the programme have an effect ( $P < x\%$ )?) (20). In cases where answers to questions about “how” the implementation of the programme led to the effects observed are needed, explanatory studies may be appropriate.

### What study design is most appropriate to meet your needs?

For studies that fall within an efficacy or effectiveness stage of evaluation, and for which there is a poor underlying evidence base on the outcomes of the digital health intervention, RCTs are the gold standard. However, in instances where the underlying evidence base linking an intervention to improved health outcomes is well established, or if it is unethical or infeasible to conduct an RCT, then a quasi-experimental or observational study design may be most realistic for collecting quantitative data. For studies that fall within an implementation science stage of evaluation, one of the three hybrid designs noted earlier in Part 4a is likely to be most appropriate. Hybrid study designs, while newly emerging in the literature, aim to differentiate between assessments of the effectiveness of a clinical strategy or intervention and the effectiveness of the service delivery or implementation strategy. The ability of a study design to measure the effectiveness overall – as well as among different subgroups of the population and different sub-components of the programme – is critical for explaining possible differences in key outcomes across study contexts. In the absence of these data, evidence on variations in effects can be difficult to understand.

### What evaluation methods are right for you?

Depending on the inferences selected, a mix of quantitative, qualitative, economic and financial evaluation activities is likely to be appropriate. Resources (human and financial), timing and audience needs will define what is feasible. Table 4.10 outlines illustrative evaluation methods by type and stage of evaluation.

**Table 4.10. Illustrative evaluation activities by type and stage of evaluation**

Formative			
		NEEDS ASSESSMENT	PROCESS EVALUATION
Stage of evaluation	Feasibility/usability	<b>Descriptive:</b> Landscape analysis, literature review <b>Qualitative:</b> Participant observation, FGDs, IDIs, case studies <b>Quantitative:</b> Mobile phone ownership and use, user profile survey <b>Mixed methods:</b> Combination of quantitative and qualitative data	<b>Qualitative:</b> FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data
	Efficacy	<b>Descriptive:</b> Contextual adaptation	<b>Qualitative:</b> FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data
	Effectiveness	<b>Descriptive:</b> Contextual adaptation	<b>Qualitative:</b> FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data
		IMPLEMENTATION MONITORING	
		<b>Qualitative:</b> Participant observation, FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data	
		<b>Qualitative:</b> Participant observation, FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data	
		<b>Qualitative:</b> Participant observation, FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data	

Formative				
NEEDS ASSESSMENT		PROCESS EVALUATION	IMPLEMENTATION MONITORING	
Stage of evaluation	Implementation science	<b>Descriptive:</b> Contextual adaptation	<b>Qualitative:</b> FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data	<b>Qualitative:</b> Participant observation, FGDs, IDIs <b>Quantitative:</b> System-generated monitoring data <b>Mixed methods:</b> Combination of quantitative and qualitative data
Summative				
OUTCOME EVALUATION		IMPACT EVALUATION	ECONOMIC EVALUATION	FINANCIAL EVALUATION
Stage of evaluation	Feasibility/usability	N/A	N/A	Affordability, financial evaluation
	Efficacy	<b>Quantitative:</b> User satisfaction, cross-sectional coverage surveys, service delivery/utilization <b>Qualitative:</b> IDIs, FGDs <b>Mixed methods:</b> Combination of quantitative and qualitative data	<b>Quantitative:</b> Disease surveillance, cross-sectional coverage or longitudinal surveys, verbal autopsies, death audits <b>Qualitative:</b> IDIs, FGDs <b>Mixed methods:</b> Convergence of quantitative and qualitative data	If comparator is available: CEA, CBA, CUA, CMA  Affordability, financial evaluation
	Effectiveness	<b>Quantitative:</b> User satisfaction, cross-sectional coverage surveys, service delivery/utilization <b>Qualitative:</b> IDIs, FGDs <b>Mixed methods:</b> Combination of quantitative and qualitative data	<b>Quantitative:</b> Disease surveillance, cross-sectional coverage or longitudinal surveys, verbal autopsies, death audits <b>Qualitative:</b> IDIs, FGDs <b>Mixed methods:</b> Combination of quantitative and qualitative data	If comparator is available: CEA, CBA, CUA, CMA  Affordability, financial evaluation
	Implementation science	<b>Quantitative:</b> User satisfaction, repeated cross-sectional coverage surveys, service delivery/utilization <b>Qualitative:</b> IDIs, FGDs <b>Mixed methods:</b> Combination of quantitative and qualitative data	<b>Quantitative:</b> Cross-sectional coverage or longitudinal surveys, verbal autopsies, death audits <b>Qualitative:</b> IDIs, FGDs <b>Mixed methods:</b> Combination of quantitative and qualitative data	If comparator is available: CEA, CBA, CUA, CMA  Affordability, financial evaluation, OneHealth modelling

CBA: cost–benefit analysis; CEA: cost–effectiveness analysis; CMA: cost-minimization analysis; CUA: cost–utility analysis; FGD: focus group discussion; IDI: in-depth interview.



## Relevant resources for evaluation

### Study designs

#### *Stepped-wedge design*

- Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol*. 2006;6:54. doi:10.1186/1471-2288-6-54.

#### *Quasi-experiment study designs*

- Harris AD, Bradham DD, Baumgarten M, Zuckerman IH, Fink JC, Perencevich EN. The use and interpretation of quasi-experimental studies in infectious diseases. *Clin Infect Dis*. 2004;38(11):1586–91.

### Qualitative methods

- Mack N, Woodsong C, MacQueen KM, Guest G, Namey E. Qualitative research methods: a data collector's field guide. Family Health International; 2005 (<https://www.fhi360.org/resource/qualitative-research-methods-data-collectors-field-guide>, accessed 13 April 2016).
- Meadows KA. So you want to do research? 3. An introduction to qualitative methods. *Br J Community Nurs*. 2003;8(10):464–9.
- Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. *Br Dent J*. 2008;204(6):291–5. doi:10.1038/bdj.2008.192.
- Burnard P, Gill P, Stewart KF, Chadwick BL. Analysing and presenting qualitative data. *Br Dent J*. 2008;204(8):429–32. doi:10.1038/sj.bdj.2008.292.
- Biddix JP. Research Rundowns [online blog]. 2009 (<https://researchrundowns.wordpress.com/>, accessed 15 April 2016).

### Quantitative methods

- Schueren F. What is a survey? 2004 (<http://www.amstat.org/sections/srms/pamphlet.pdf>, accessed 13 April 2016).
- Sonenstein F. Issues in survey research design (Course Number 380.711.01). Johns Hopkins School of Public Health Open Courseware. 2012 (<http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/surveyresearchdesign/coursePage/index/>, accessed 13 April 2016).
- Johnson TP. Handbook of health survey methods. John Wiley & Sons; 2015 (<http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118002326,subjectCd-HE55.html>).
- Meadows KA. So you want to do research? 4: An introduction to quantitative methods. *Br J Community Nurs*. 2003;8(11):519–26.
- Meadows KA. So you want to do research? 5: Questionnaire design. *Br J Community Nurs*. 2003;8(12):562–70.
- Zurovac D, Otieno G, Kigen S, Mbithi AM, Muturi A, Snow RW, Nyandigisi A. Ownership and use of mobile phones among health workers, caregivers of sick children and adult patients in Kenya: cross-sectional national survey. *Global Health*. 2013;9:20. doi:10.1186/1744-8603-9-20.
- Crawford J, Larsen-Cooper E, Jezmana Z, Cunningham SC, Bancroft E. SMS versus voice messaging to deliver MNCH communication in rural Malawi: assessment of delivery success and user experience. *Glob Health Sci Pract*. 2014;2(1):35–46. doi:10.9745/GHSP-D-13-00155.

### Mixed methods

- Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs – principles and practices. *Health Serv Res*. 2013;48(6 Pt 2):2134–56. doi:10.1111/1475-6773.12117.



- Gilson L, Hanson K, Sheikh K, Akua Agyepong I, Ssengooba F, Bennett S. Building the field of health policy and systems research: social science matters. *PLoS Med.* 2011;8(8):e1001079. doi:10.1371/journal.pmed.1001079.

### **Participatory action research (PAR)**

- Smith L, Rosenzweig L, Schmidt M. Best practices in the reporting of participatory action research: embracing both the forest and the trees. *Counseling Psychologist.* 2010;38(8):1115–38.
- Israel BA, Schulz AJ, Parker EA, Becker AB, Allen AJ, III, Guzman JR. Chapter 3: Critical issues in developing and following community-based participatory research principles. In: Minkler M, Wallerstein N, editors. *Community-based participatory research for health: from processes to outcomes*, second edition. Wiley; 2008.

## **References**

1. WHO evaluation practice handbook. Geneva: World Health Organization; 2013 ([http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf), accessed 29 April 2016).
2. Pritchett L, Samji S, Hammer J. It's all about MeE: using structured experiential learning ("e") to crawl the design space. Washington (DC): Center for Global Development; 2013.
3. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *Methods for the economic evaluation of health care programmes*. New York (NY): Oxford University Press; 2005.
4. Gartlehner G, Hansen R, Nissman D, Lohr K, Carey T. *Criteria for distinguishing effectiveness from efficacy trials in systematic reviews*. Rockville (MD): Agency for Healthcare Research and Quality; 2006.
5. Glossary of selected planning, monitoring and evaluation terms. Measurement, Learning and Evaluation (MLE) Project, Urban Reproductive Health Initiative; 2013 (<https://www.urbanreproductivehealth.org/toolkits/measuring-success/glossary-selected-planning-monitoring-and-evaluation-terms>, accessed 4 May 2016).
6. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. *BMJ.* 2013;347:f6753. doi:10.1136/bmj.f6753.
7. *Building our understanding: key concepts of evaluation*. Atlanta (GA): United States Centers for Disease Control and Prevention, Health Communities Program; undated ([http://www.cdc.gov/nccdphp/dch/programs/healthycommunitiesprogram/tools/pdf/apply\\_theory.pdf](http://www.cdc.gov/nccdphp/dch/programs/healthycommunitiesprogram/tools/pdf/apply_theory.pdf), accessed 13 May 2016).
8. Song JW, Chung KC. Observational studies: cohort and case-control studies. *Plast Reconstr Surg.* 2010;126(6):2234–42.
9. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health.* 2002;56(2):119–27.
10. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care.* 2012;50(3):217–26.
11. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol.* 2006;6:54. doi:10.1186/1471-2288-6-54.
12. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet.* 2002;359(9301):145–9.
13. Last, LM. *A dictionary of epidemiology*, second edition. New York (NY): Oxford University Press; 1988.
14. Gordis L. *Epidemiology*, fifth edition. Elsevier; 2014 (<http://store.elsevier.com/Epidemiology/Leon-Gordis/isbn-9781455737338/>, accessed 29 April 2016).
15. Donalek JG. Choosing among qualitative traditions. *Urol Nurs.* 2004;24(5):409–10, 422.

16. Lindquist S. Qualitative research designs. University of Missouri – St Louis; undated (<http://www.umsl.edu/~lindquists/qualdsgn.html>, accessed 6 May 2016).
17. Neill J. Analysis of professional literature. Class 6: Qualitative research I. Outdoor Education Research & Evaluation Center; 2006 (<http://wilderdom.com/OECourses/PROFLIT/Class6Qualitative1.htm>, accessed 6 May 2016).
18. Harris AD, Bradham DD, Baumgarten M, Zuckerman IH, Fink JC, Perencevich EN. The use and interpretation of quasi-experimental studies in infectious diseases. *Clin Infect Dis*. 2004;38(11):1586–91.
19. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J*. 2003;20(1):54–60.
20. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol*. 1999;28(1):10–8.
21. Peters D, Tran N, Adam T. Implementation research in health: a practical guide. Geneva: Alliance for Health Policy and Systems Research, World Health Organization; 2013.
22. Meadows KA. So you want to do research? 3. An introduction to qualitative methods. *Br J Community Nurs*. 2003;8(10):464–9.
23. Appraising scientific evidence: qualitative versus quantitative research. The Association of Faculties of Medicine of Canada; undated (<http://phprimer.afmc.ca/Part2-MethodsStudyingHealth/Chapter5AssessingEvidenceAndInformation/Appraisingscientificevidencequalitativeversusquantitativeresearch>, accessed 6 May 2016).
24. Creswell JW. Research design: qualitative, quantitative and mixed methods approaches, second edition. SAGE Publications; 2003.
25. Qualitative versus quantitative research. Xavier University Library; 2012 ([http://www.xavier.edu/library/students/documents/qualitative\\_quantitative.pdf](http://www.xavier.edu/library/students/documents/qualitative_quantitative.pdf), accessed 6 May 2016).
26. Mack N, Woodsong C, MacQueen KM, Guest G, Namey E. Qualitative research methods: a data collector's field guide. Research Triangle Park (NC): Family Health International; 2005 (<https://www.fhi360.org/resource/qualitative-research-methods-data-collectors-field-guide>, accessed 6 May 2016).
27. Stewart K, Gill P, Chadwick B, Treasure E. Qualitative research in dentistry. *Br Dent J*. 2008;204(5):235–9.
28. Meadows KA. So you want to do research? 4: An introduction to quantitative methods. *Br J Community Nurs*. 2003;8(11):519–26.
29. Schueren F. What is a survey? 2004 (<http://www.amstat.org/sections/srms/pamphlet.pdf>, accessed 6 May 2016).
30. Issues in survey research design, Spring 2012. In: Johns Hopkins School of Public Health Open Courseware [website]. Johns Hopkins Bloomberg School of Public Health; 2011 (<http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/surveyresearchdesign/coursePage/index/>, accessed 6 May 2016).
31. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs – principles and practices. *Health Serv Res*. 2013;48(6 Pt 2):2134–56.
32. Qualitative research: getting started [website]. Durham (NC): Duke University Libraries; 2015 (<http://guides.library.duke.edu/qualitative-research>, accessed 6 May 2016).
33. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. *Br Dent J*. 2008;204(6):291–5.
34. Israel BA, Schulz AJ, Parker EA, Becker AB, Allen AJ, III, Guzman RJ. Community-based participatory research for health: from processes to outcomes (Chapter 3). In: Minkler M, Wallerstein N, editors. Critical issues in developing and following CBPR principles, second edition. Wiley; 2010.

35. Smith L, Rosenzweig L, Schmidt M. Best practices in the reporting of participatory action research: embracing both the forest and the trees. *Counseling Psychologist*. 2010;38(8):1115–38.
36. Burnard P, Gill P, Stewart K, Treasure E, Chadwick B. Analysing and presenting qualitative data. *Br Dent J*. 2008;204(8):429–32.
37. Meadows KA. So you want to do research? 5: Questionnaire design. *Br J Community Nurs*. 2003;8(12):562–70.
38. Kadam P, Bhalerao S. Sample size calculation. *Int J Ayurveda Res*. 2010;1(1):55–7.
39. Meadows KA. So you want to do research? 1: an overview of the research process. *Br J Community Nurs*. 2003;8(8):369–75.
40. Probability sampling. In: Research methods knowledge base [website]. William M.K. Trochim; 2006 (<http://www.socialresearchmethods.net/kb/sampprob.php>, accessed 6 May 2016).
41. Meadows KA. So you want to do research? 6: Reporting research. *Br J Community Nurs*. 2004;9(1):37–41.
42. Gilson L, Hanson K, Sheikh K, Agyepong IA, Ssengooba F, Bennett S. Building the field of health policy and systems research: social science matters. *PLoS Med*. 2011;8(8):e1001079.
43. Public sector business cases: using the five case model. Green Book supplementary guidance on delivering public value from spending proposals. London: HM Treasury, United Kingdom Government; 2003.
44. Thokala P, Dixon S, Jahn B. Resource modelling: the missing piece of the HTA jigsaw? *Pharmacoeconomics*. 2015:1–11.
45. Sullivan SD, Mauskopf JA, Augustovski F, Caro JJ, Lee KM, Minchin M et al. Budget impact analysis – principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. *Value Health*. 2014;17(1):5–14. doi:10.1016/j.jval.2013.08.2291.
46. The Green Book: appraisal and evaluation in central government. London: HM Treasury, United Kingdom Government; 2003.
47. Kumar S, Williams AC, Sandy JR. How do we evaluate the economics of health care? *European J Orthodontics*. 2006;28(6):513–9.
48. Cunningham SJ. An introduction to economic evaluation of health care. *J Orthod*. 2001;28(3):246–50.
49. O'Brien K, Kay L, Fox D, Mandall N. Assessing oral health outcomes for orthodontics – measuring health status and quality of life. *Community Dent Health*. 1998;15(1):22–6.
50. Robinson R. Cost-benefit analysis. *BMJ*. 1993;307(6909):924.
51. Dobrev A, Jones T, Kersting A, Artmann J, Streotmann K, Stroetmann V. Methodology for evaluating the socio-economic impact of interoperable EHR and ePrescribing systems. Bonn: European Commission; 2008 ([http://www.ehr-impact.eu/downloads/documents/EHRI\\_D1\\_3\\_Evaluation\\_Methodology\\_v1\\_0.pdf](http://www.ehr-impact.eu/downloads/documents/EHRI_D1_3_Evaluation_Methodology_v1_0.pdf), accessed 13 May 2016).
52. The economic impact of eHealth: method, case studies, summary results. eHealth Impact; undated (<http://www.ehealth-impact.org/>, accessed 6 May 2016).
53. OneHealth Tool. In: Cost effectiveness and strategic planning (WHO-CHOICE) [website]. Geneva: World Health Organization; 2016 (<http://www.who.int/choice/onehealthtool/en/>, accessed 6 May 2016).



## **Chapter 5: Assessing data sources and quality for M&E**



## HOW WILL THIS CHAPTER HELP ME?

This chapter will:

- ✓ Help you to assess whether existing intervention<sup>5</sup> data are aligned with the claims/goals that the digital health intervention hopes to achieve.
- ✓ Show you how to map the flow of data at each level of the health system to understand the key stakeholders who support data collection and use.
- ✓ Guide you in assessing the process of data collection, storage, analytics/dashboards, management and use, in order to identify and address issues of data quality.
- ✓ Provide recommendations for improving the overall data-related processes to strengthen the intervention's ability to validate its claims using sound evidence.

**F**urther to articulating your monitoring and evaluation (M&E) goals and designing the plans for undertaking the M&E activities, another key component is determining the data requirements. Implementers need to determine whether their project will yield the right type of data of sufficient quantity and quality to support the claims articulated by the project.

Assessing digital data helps to identify and bridge gaps between what is intended to be measured and the practical ability to measure the variables that will show whether or not the intended impact is occurring. Part 5a will introduce an approach to conducting this data quality assessment, Part 5b provides an empty worksheet with instructions, and Part 5c presents a sample application of steps 1 and 2.

### Why assess your digital data?

If you wonder . . .

- Will we be able to prove that our intervention works the way that we think it does based on the data?
- Are we collecting the right data for our intervention?
- What steps can we take to improve our data quality?
- Are there additional opportunities for data collection and use?

<sup>5</sup> "Intervention" in this Guide can also refer to projects, programmes, initiatives and other activities that are being monitored and evaluated.



# Part 5a: Introducing the data quality assessment approach and how to do it

## Key components

Through a process of claims formulation and a review of available data sources and data quality, this approach will help to identify whether a claim can be supported by existing data and, if not, to identify alternative claims that may be better aligned with the programme design and available data streams. If applied early in the programme planning phase, it can be used to strengthen data sources to improve the implementer's ability to support their intervention's claims with robust data. Figure 5.1 presents the key components of data quality assessment.

**Step 1** walks the user through identifying their main intervention claims, listing associated indicators and data sources, and assessing whether the proposed claims can feasibly be backed up by the available data. For more details on defining intervention claims, please refer to Chapter 2, Part 2a.

**Step 2** helps the implementation team develop a visual depiction of how the data flows across the organization. This illustration can help identify gaps and opportunities in data collection, analysis and use (2).

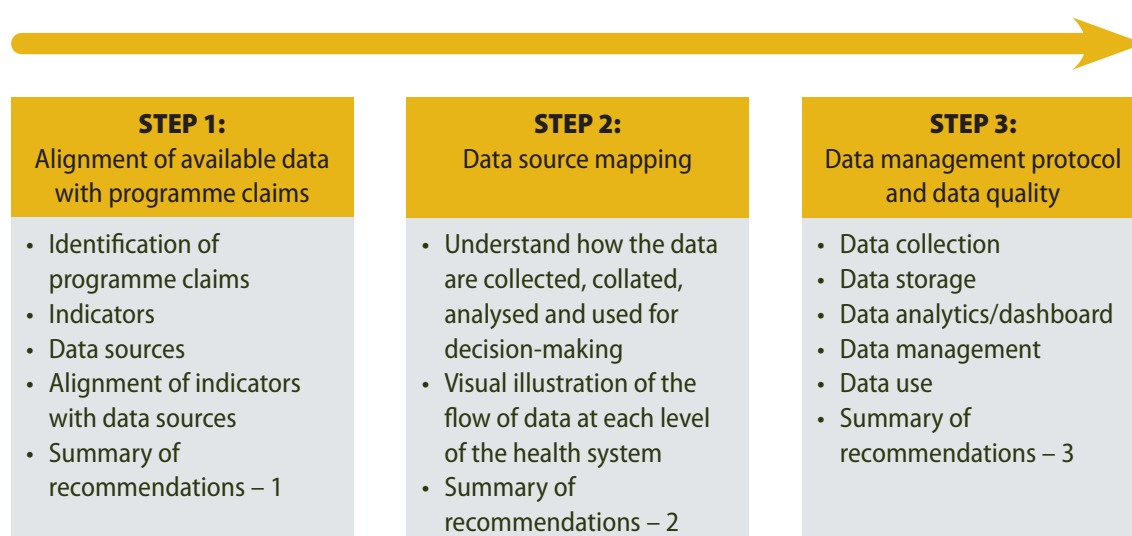
**Step 3** delves into the nuts and bolts of processes that can result in either good or poor data quality, highlighting best practices at each of the five stages of data (i.e. collection, storage, analytics, management and use). The approach for assessing data quality and use has been adapted from the PRISM Framework, which is widely accepted and used to assess routine health information system (RHIS) performance (3).



**Claim:** A statement of anticipated benefits of the digital health system or intervention.

**Indicator:** A “quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor” (1).

**Figure 5.1. Key components of the three steps of data quality assessment**





## How should data quality assessment be used?

If the organization implementing the intervention has someone with M&E expertise on its team, data assessment should be used by the team internally under the leadership of the M&E expert. Step 2 would benefit from bringing together the key people who are involved in collecting, analysing and using the data to discuss the data flow process and identify ways of collaborating to improve the process. Parts of Step 3 that focus on adequate training and organizational support for data use are somewhat subjective, and would benefit from the opinions of the intervention project's broader implementation and leadership team.

If the implementing organization does not have M&E expertise on its team, we recommend hiring an external M&E consultant to work with the team to apply the tool.

Assuming that intervention documentation – including a list of the goals – and information on data sources are readily available, this tool should not take more than a total of three hours to apply.

## We assessed the quality of our data. Now what?

A data quality assessment can only aid in improving intervention data processes if it is followed up with action and repeat assessments. It can be used at different stages of the intervention implementation cycle to help plan, improve and assess the data processes. Each step of the assessment yields a “summary of recommendations”. Once completed, the project team should discuss these summaries and define specific action points and responsibilities. Depending on the breadth of the gaps identified, the project data processes can be reassessed using data quality assessment every year.

# Part 5b: Data quality assessment worksheet and instructions

This worksheet can be printed and used by the team to walk through the process.

## Data quality assessment worksheet

<b>Name of the project:</b>	<b>Project personnel interviewed:</b>  <b>Name:</b>  <b>Title:</b>
<b>Name of the assessor:</b>	
<b>Date of assessment:</b>	

## Step 1: Assessment of alignment of available data with programme claims

### Step 1.i. Identification of the intervention's claims

Claims are statements of anticipated benefits of the digital health system or intervention. A claim may be thought of as the expected “result” of the digital health intervention.

**Instructions:** This section should be completed in collaboration with the implementers of the digital health intervention and/or through a review of the project documents. List 5–10 key claims that the digital health intervention aims to achieve or reports in its standard documents. For example, “Increase in the utilization of antenatal care (ANC) will be greater in intervention areas than in non-intervention areas”. If intervention claims are not identifiable, please refer to Chapter 2, Part 2a, on articulating claims.

No.	Key intervention claims
1	
2	
3	
4	
5	

### Step 1.ii. Indicators

**What is an indicator?** An indicator is a clearly defined metric used to measure a particular intervention output that is intended to support one or more of the claims, which are listed in Step 1.i. For more details on how to develop “SMART” indicators and resources on standardized indicator repositories, please refer to Chapter 2, Part 2c.

**Instructions:** For each claim listed in Step 1.i, identify 1–5 specific **indicators** that would support the claim. If a claim is already in the form of a specific indicator, it can be copied directly from the list in Step 1.i.

Identify the **numerator (N)** and **denominator (D)** for each of the indicators. For each N and D, indicate the beginning and end month and year for which the data are available (i.e. the time frame). For the most part, each member of the numerator group (e.g. women accessing timely ANC) should come from the denominator group (e.g. all participating pregnant women).

For indicators that measure specific health outcomes, ensure, as far as possible, alignment with global standard indicators defined by WHO, other UN agencies or other normative bodies (e.g. improvements in ANC should be measured as the

percentage of women of reproductive age who receive any ANC or at least four ANC contacts). Recently, the Global Strategy for Women's, Children's and Adolescents' Health developed high-level "standardized" indicator sets, some of which may be applicable to the project. In other cases, there may be publicly available *national* indicators, which should be used to ensure optimal comparability of the data within that country or with international standards.

If no denominator applies (e.g. change in number of participants over time), then the time period and/or geographic area should be included as part of the indicator (e.g. 10% increase in the number of subscribers in three specified districts or villages). Indicators of this type are considered to be suboptimal indicators by many, as it is difficult to appreciate the magnitude of impact without a denominator as a frame of reference.

Claim No.	Indicator	Numerator (N)		Time frame (from – to)	Denominator (D)		Time frame (from – to)
		N1			D1		
		N2			D2		
		N3			D3		
		N4			D4		
		N5			D5		

### Step 1.iii. Data sources

**What is a data source?** Typically, this refers to information collected systematically (purposively) to monitor the progress of an intervention or to measure the magnitude of, or changes in, an outcome. However, with digital data systems, a wealth of data may be generated by the system, which are often underutilized and sometimes discarded or not stored. These data include what is known as "meta-data", or digital data implicitly collected by digital systems, such as timestamps (indicating *when* a specific datum was collected), or geolocation (indicating *where* a datum was collected). System-generated data combined with purposefully collected data offer a unique combination of information that can be used to monitor implementation fidelity as well as to measure changes in system/user performance over time.

**Instructions:** List all the sources of data (including system-generated data) you are collecting.

**System-generated or survey data:** For each data source, indicate whether the data are directly generated during routine operations/health-care delivery, or collected as part of a survey/study. System-generated/routine data refers to data that are collected on a continuous basis as part of routine operations. Survey data refers to data that are collected cross-sectionally or as a part of a specific one-time-only study.

**Data format:** Indicate whether the data are available for use in paper or digital format. If the intervention collects data on paper and then digitizes them, please indicate the format in which the data are currently available. If data are available in both paper and digital format, indicate digital.

**Frequency of data collection:** For each data source, write in words how frequently the data are collected.

**Comparison group:** For each data source, identify whether similar data are available for a comparison group (Y) or not (N).

**Indicators:** From Step 1.ii, identify all the numerators (N) and denominators (D) that can be determined from each of the data sources. List them in the "Indicators No." column on the right, using their N/D numbers from the form in Step 1.ii.

Data sources for intervention and comparison arms may be listed separately. If it is planned that comparison data are to be abstracted from a secondary data source (e.g. national surveys, DHS, MICS), please list those data sources as well.

Name of data source	System-generated OR survey data	Data format: paper OR digital	Frequency of collection	Comparison group (Y/N)	Indicators No.
E.g. Women's household survey	Collected using routine mobile phones service delivery		Baseline	Y	N1, N3, D4, D10
E.g. District health-care facility records	Survey data		At each follow-up visit	N	N2, N5, N6, D1–D3

#### Step 1.iv. Alignment of indicators with data sources

**Do I have the necessary information, given my data streams, to support my claims?** This section brings together the work done in the three sections above. We've identified the claims the intervention would like to make with regard to the outcomes and impact of the digital health intervention, we've described the indicators that would be needed to support such a claim, and we've examined the breadth and depth of the data available to populate these indicators. Putting these together, let's review the claims and examine whether numerator and denominator data are available to support these claims.

This section aims to assess whether sufficient data are available to support one or more claims identified in Step 1.i, based on the information gathered in Steps 1.ii and 1.iii.

**Instructions:** Please summarize the data below to identify the claims that have data to support them.

Claim No. (see Step 1.i)	Indicator (see Step 1.ii)	Numerator (see Step 1.ii)	Data available (Y/N)? (see Step 1.iii)	Denominator (see Step 1.ii)	Data available (Y/N)? (see Step 1.iii)
		N1		D1	
		N2		D2	
		N3		D3	
		N4		D4	
		N5		D5	

### Summary of recommendations – Step 1

Based on review of the data available, the M&E team may make the following recommendations:

#### If sufficient data are available to support one or more claims:

- Summarize the claims that can be made, given the available data.
- Generate recommendations for data collection for any claims that do not have data to support them.
- Identify any additional claims that might be possible, given the data.

#### If sufficient data are not available to support at least one claim:

- Generate recommendations for data that should be collected to support the claims. Recommend potential data collection methods/sources that might facilitate such data collection.
- Identify additional claims that might be possible, given the data.

**Irrespective of the results of Step 1, proceed to Step 2.**

## Step 2: Data mapping

### Step 2.i. Understand how the data are collected, collated, analysed and used for decision-making

This understanding is enhanced by completing the chart in the next step, which allows us to identify deficiencies in the system at any point from data collection to data use – deficiencies that can indirectly affect data quality (3).

### Step 2.ii. Visual illustration of the flow of data at each level of the health system

Completing the chart on the next page will illustrate all the stakeholders that are involved in gathering the data and will thus help to visualize the flow of data at each level of the health system.

**Instructions:**<sup>6</sup> Use the chart on the next page to illustrate the flow of information at each level of the health system.

- Each row represents a level of data collection or group of stakeholders, such as community health workers, local health-care facilities, district health-care facilities. Modify the number and content of the rows as needed, to make the chart relevant to the project.
- Each column represents a stage in the life-cycle of the data. Modify the columns as needed, to make the chart relevant to the processes in your intervention from data collection to data use.
- In each cell in the chart, indicate if and how the stakeholder listed in the row conducts the activity in the column. The information added should include:
  - ✓ What data are collected?
  - ✓ Are the data in paper or digital format?
  - ✓ How often is the activity conducted?
  - ✓ For storage: Where are the data stored?
  - ✓ For analysis: At what level is the analysis conducted?
  - ✓ For use: For what purpose are the data used?
- Draw arrows between cells to illustrate transfer of information from one level to the next.
- For an illustrative application of the data flow map, refer to Part 5c.
- If data collection mechanisms are different for the intervention and comparison arms of a project, two separate maps can be made.

<sup>6</sup> More detailed instructions on the application of this tool can be accessed here: [https://www.k4health.org/sites/default/files/migrated\\_toolkit\\_files/DDIU\\_Information\\_Use\\_Mapping.pdf](https://www.k4health.org/sites/default/files/migrated_toolkit_files/DDIU_Information_Use_Mapping.pdf)

Stakeholder	Data collection	Compilation	Storage	Analysis	Reporting	Use
Research worker						
Client-generated						
Front-line health worker						
Private clinic						
NGO						
Government facility						
District level						
Regional level						
National level						

### Summary of recommendations – Step 2

The Data Map allows the M&E team to outline a logical sequence of steps needed to improve the flow of data at each level of the health system. However, the component steps of Step 2 outlined above are not absolute. Generation of recommendations should be guided by what is reasonable for each project.

Some questions that can help to guide this process include:

- Are there any stakeholders that are currently not part of the system, who should be part of the system?
- Are there opportunities in the system for facilities/organizations/other stakeholders to analyse their own data?
- Are there additional gaps/opportunities for data use at any level of the health system?
- Are any changes in the frequency of data collection/analysis/reporting/use recommended at any level of the health system?
- Would any changes in the level of data disaggregation be appropriate to improve data use?
- Are there any other pertinent challenges that threaten data quality?

Based on the identified gaps and opportunities, a revised map should be developed with additional opportunities for data analysis, reporting and use, and including recommended changes in the frequency of collection and/or in the level of disaggregation of data.



## Step 3: Assessment of data management protocols and data quality

**Instructions:** The following section refers to the data that are required to measure the indicators identified in Step 2. The M&E team should review each of the items in the checklist to generate summary recommendations.

Assessment items	Response categories		
	Yes	No	Inconsistent
<b>STEP 3.i: DATA COLLECTION</b>			
<b>Item 1: Data elements</b>			
1.1. Can the relevant databases/data sources be linked via a common <b>client identifier</b> ?			
1.1. For data collection using digital devices, are there built-in rules to validate the <b>identifier</b> in each database? (Are there rules or procedures in place to prevent identifier errors in the data?)			
1.3. Is searching for records by <b>identifier</b> possible?			
1.4. For each of the indicators identified above:			
a. Are the <b>data element</b> names clearly defined in the questionnaire/form?			
b. Have these <b>data elements</b> been previously tested?			
c. Are appropriate parameters (range, format) and validation checks defined for the <b>data elements</b> ?			
d. Is it possible to incorporate international standard definitions of the <b>data elements</b> into the system (if not already done)?			
1.5. Have protocols been established to identify missing data, and are they being followed?			
1.6. Have the staff been trained in data entry (either paper or digital)?			
<b>Item 2: Data collection using paper forms</b>			
2.1. Are paper forms being used to capture some of the data?			
2.2. Are paper forms being manually entered into a digital system? If so:			
a. Have the staff been trained in data collection using paper forms?			
b. Is there an identifier to link the paper data to the digital data?			
c. Are random checks and other quality control mechanisms in place to assess accuracy of data-entry procedures?			
<b>STEP 3.ii: DATA STORAGE</b>			
<b>Item 3: Data transmission</b>			
3.1. Is there a system for data backup and archiving?			
3.2. Is there ability to update/edit client records manually?			
3.3. Are efforts taken to minimize the delay between data collection and transmission?			
3.4. Is there a system in place to ensure accountability between data collected in the field and data received in the server? (e.g. many organizations use “digital tokens” to confirm that a digital report has been successfully sent and received, to help understand where data might be going missing).			
<b>Item 4: Data security</b>			
4.1. Is client information encrypted?			
4.2. Are measures in place to limit access to the data at each stage of collection and storage?			
<b>STEP 3.iii: DATA ANALYTICS AND DASHBOARDS</b>			
<b>Item 5: Data dashboard</b>			
5.1. Is there ability to export data for analysis?			
5.2. Does a data dashboard exist to easily download data?			
5.3. Does a data dashboard exist to produce the following:			
a. Calculate each of the indicators identified under Step 1.ii?			
b. Data summary (aggregated data) at community/district/other unit level?			
c. Monthly/quarterly reports?			
d. Comparisons of data over time?			
5.4. Do you think that the monthly/quarterly report is complex and difficult to follow?			
5.5. Are the data available to leaders/health officials at each level of the organization?			

Assessment items	Response categories		
	Yes	No	Inconsistent
5.6. How frequently are new data entered by individuals at each level of the organization?	1 – Daily 2 – Weekly 3 – Monthly 4 – Quarterly 5 – Annually 6 – Uncertain		
5.7. How frequently are data accessed by individuals at each level of the organization?	1 – Daily 2 – Weekly 3 – Monthly 4 – Quarterly 5 – Annually 6 – Uncertain		
5.8. Are visual forms of the data (e.g. graphs, charts, tables, maps) readily available and/or easy to generate?			
<b>STEP 3.iv: DATA MANAGEMENT</b>			
<b>Item 6: Staff training</b>			
6.1. In total, how many staff members are involved with data collection for this specific project?	No.:		
6.2. How many days in the last year, on average, have these staff members been trained on data collection procedures for this project?	Days:		
6.3. Is there a plan for refresher training in data collection?			
<b>Item 7: Supervision</b>			
7.1. Has the role of supervision of the data collection process been assigned (to one or multiple individuals)?			
7.2. How often does each data collection staff member receive an oversight visit?	0 – No schedule 1 – Weekly 2 – 2 weekly 3 – Monthly 4 – Quarterly 5 – Annually		
7.3. Is the assessment from the supervision visits documented?			
7.4. Has the role of following up on the action steps after supervision been assigned?			
<b>Item 8: Data documentation</b>			
8.1. Does documentation (e.g. a written protocol) exist to describe the use of the mobile/paper tools for data collection?			
8.2. Does documentation exist for management of the data, roles and responsibilities?			
8.3. Are skip patterns in data clearly documented?			
8.4. Is there a documented system to check for data accuracy (such as random checks, regular tabulations to assess that the data make sense)?			
<b>STEP 3.v: DATA USE</b>			
<b>Item 9: Regularity of meetings</b>			
9.1. Are meetings conducted at the organizational and/or district level to review managerial and administrative data concerns?			
9.2. How frequently do these meetings take place?	0 – No schedule 1 – Weekly 2 – 2 weekly 3 – Monthly 4 – Quarterly		
9.3. How many times did such a meeting take place in the last three months?			
9.4. Is an official record of these meetings maintained?			
9.5. If yes, please check the meeting records for last three months to identify whether the following topics were discussed:			
a. Management of the data, including data quality, reporting and timeliness of reporting			
b. Discussion on the utilization of patient record data to inform services			
c. Any follow-up decisions based on the above discussions.			

### Summary of recommendations – Step 3

This section provides a checklist of items for the M&E team to assess within the existing system of data quality assurance and management, as well as relevant questions for each item. While the responses to the questions will help to identify areas for improvement, specific recommendations will depend on ensuring the evaluation of the items that are most pertinent to the project, and will be shaped by the M&E team's expertise in the subject matter.

#### Step 3.i. Data collection

##### *Item 1: Data elements*

- Are data elements for each of the indicators clearly defined? If not, what changes are recommended?
- Are there automated or manual validation checks for the data elements in the data set? If not, state the recommended changes to improve quality checks.
- Is it possible to link relevant databases to support the project's/programme's claims? Which databases should be linked and what identifiers might be used to link them?

##### *Item 2: Data collection using paper forms*

- If data from paper forms are needed to support the claims, is it feasible to readily access the data and is there a system linking the paper data to digital data? If not, what recommendations can be made to improve data use and accessibility?

#### Step 3.ii. Data storage

##### *Item 3: Data transmission*

- Is the system for data back-up, archiving and updating adequate?

##### *Item 4: Data security*

- Are adequate data security measures in place?

#### Step 3.iii. Data analytics/dashboards

##### *Item 5: Data dashboard*

- Is there a convenient way to export the data into a usable format and share it?
- For each indicator, is the level of disaggregation and reporting in alignment with how the indicator should be measured? If not, what changes should be made?
- Is the dashboard intuitive to use and do the leadership report using it?

#### Step 3.iv. Data management

##### *Item 6: Staff training*

- Given the size of the project, are there enough staff to support it?
- Given the complexity of the project, has sufficient training time been built in?

##### *Item 7: Supervision*

- If a supervisory role exists, does the level of supervision appear to be adequate given the scale of the operation?

##### *Item 8: Data documentation*

- Does documentation exist for key programme operations and procedures? Identify areas that would benefit from documentation.

#### Step 3.v. Data use

- Are meetings held regularly to discuss issues of data management and use?
- Based on existing meeting records, is data quality discussed during these meetings?

## Part 5c: Sample application of data quality assessment

This part applies the data quality assessment procedures with a sample programme of the Kenya Medical Research Institute (KEMRI)'s "Texting to improve testing" (TextIT) project. The TextIT project provides theory-based individually tailored text-messaging intervention aimed at improving the retention of mothers and babies in programmes for elimination of mother-to-child transmission (EMTCT) of HIV.

### Step 1. Alignment of available data with programme claims

#### Step 1.i. Identification of programme claims

	Key claims
1	A greater proportion of women at health-care facilities implementing TextIT will attend clinic within 8 weeks postpartum compared with women at health-care facilities implementing standard care
2	Infants of women at health-care facilities implementing TextIT will be more likely to have virological HIV testing compared with infants of women at health-care facilities implementing standard care

#### Step 1.ii. Indicators

Claim No.	Indicator	Numerator (N)		Time frame (from – to)	Denominator (D)		Time frame (from – to)
1	Proportion of pregnant women who attend postnatal clinic within 8 weeks postpartum in intervention arm	N1	Number of women who attended the health-care facility postpartum at all health-care facilities in intervention arm	Jan 2015–Dec 2015	D1	Total number of pregnant women who attended the postnatal clinic at baseline at all health-care facilities in intervention arm	Jan 2015–Dec 2015
1	Proportion of pregnant women who attend postnatal clinic within 8 weeks postpartum in the control arm	N2	Number of women who attended the health-care facility postpartum at all health-care facilities in the control arm	Jan 2015–Dec 2015	D2	Total number of pregnant women who attended the postnatal clinic at baseline at all health-care facilities in the control arm	Jan 2015–Dec 2015
2	Proportion of infants receiving viral testing within 8 weeks postnatal in intervention arm	N3	Number of infants who received viral testing within 8 weeks postnatal in intervention arm	Jan 2015–Dec 2015	D3	Total number of live infants or total number of pregnancies registered at baseline in intervention arm	Jan 2015–Dec 2015
2	Proportion of infants receiving viral testing within 8 weeks postnatal in control arm	N4	Number of infants who received viral testing within 8 weeks postnatal in control arm	Jan 2015–Dec 2015	D4	Total number of live infants or total number of pregnancies registered at baseline in control arm	Jan 2015–Dec 2015

### Step 1.iii. Data sources

Name of data source	System-generated OR survey data	Data format: paper OR digital	Frequency of collection	Comparison group (Y/N)	Indicators No.
Baseline survey	System-generated data (abstracted from patient charts, antenatal care [ANC] register, mother–baby booklet)	Paper/digital	Once/woman	Y	D1, D2
Postpartum-1 survey (mother)	System-generated data (abstracted from patient charts, postnatal care [PNC] register, HIV-exposed infant [HEI] register, mother–baby booklet)	Paper/digital	Once/woman	Y	N1, N2, D3, D4
Postpartum-2 survey (baby)	System-generated data (abstracted from patient charts, HEI register, mother–baby booklet)	Paper/digital	Once/woman	Y	N3, N4, D5, D6

### Step 1.iv. Alignment of indicators with data sources

Claim	Numerator (N)	Data available (Y/N)?	Denominator (D)	Data available (Y/N)?
1	N1	Y	D1	Y
1	N2	Y	D2	Y
2	N3	Y	D3	Y
2	N4	Y	D4	Y

## Step 2. Data mapping

### 1. Current data map

	Data collection	Compilation	Storage	Analysis	Reporting	Use
<b>Mentor mother</b>	Abstract data from registers on paper, client interviews (demographic info)	Abstracted data are pooled together in ODK	In paper forms and in ODK			
<b>Study team (KEMRI)</b>				Data analysed		
<b>International donors</b>					Reported to WHO, publication issued	
<b>Kenya MOH</b>					Reported to National AIDS and STI Control Program (NASCOP), county MOH	Use in planning national programmes and policies

MOH: ministry of health; ODK: Open Data Kit; STI: sexually transmitted infection.

## 2. Map with additional opportunities for data use (additions in red)

	Data collection	Compilation	Storage	Analysis	Reporting	Use
<b>Mentor mother</b>	Abstract data from registers on paper, client interviews (demographic information)	Abstracted data are pooled together in ODK	In paper forms and in ODK		Reported to mentor mothers to understand progress	Use in improving quality and coverage of care
<b>Study team (KEMRI)</b>				Data analysed		
<b>International donors</b>					Reported to WHO, publication issued	
<b>Kenya MOH</b>					Reported to National AIDS and STI Control Program (NASCO), county MOH	Use in planning national programmes and policies

MOH: ministry of health; ODK: Open Data Kit; STI: sexually transmitted infection.



### FURTHER READING

Additional information on how to use the data flow mapping tool is available at:

[https://www.k4health.org/sites/default/files/migrated\\_toolkit\\_files/DDIU\\_Information\\_Use\\_Mapping.pdf](https://www.k4health.org/sites/default/files/migrated_toolkit_files/DDIU_Information_Use_Mapping.pdf)

PRISM framework and other tools for strengthening routine health information systems (RHIS) performance are available at:

<http://www.cpc.unc.edu/measure/resources/publications/wp-13-138>



## References

1. WHO evaluation practice handbook. Geneva: World Health Organization; 2013 ([http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf), accessed 5 May 2016).
2. Information use mapping: an assessment tool for identifying opportunities for feedback mechanisms and data use. Data demand and information use. Part Two: Strategies and tools. MEASURE Evaluation. Washington (DC): United States Agency for International Development (USAID); undated ([https://www.k4health.org/sites/default/files/migrated\\_toolkit\\_files/DDIU\\_Information\\_Use\\_Mapping.pdf](https://www.k4health.org/sites/default/files/migrated_toolkit_files/DDIU_Information_Use_Mapping.pdf), accessed 24 March 2016).
3. Tools for data demand and use in the health sector. MEASURE Evaluation manual. Washington (DC): United States Agency for International Development (USAID); 2011 (<http://www.cpc.unc.edu/measure/resources/publications/ms-11-46-d>, accessed 24 March 2016).





## Chapter 6: Reporting your findings: the mHealth Evidence Reporting and Assessment (mERA) checklist<sup>7</sup>

<sup>7</sup> The material in this chapter has been adapted from the mERA checklist, which was developed by a group of experts assembled as part of the WHO mHealth Technical Evidence Review Group (mTERG). Contributors outside of mTERG were recruited through professional and academic networks. For the complete article on mERA, please see Agarwal et al., 2016 (1).



## HOW WILL THIS SECTION HELP ME?

This section will:

- ✓ Provide concrete guidance on how to report your digital health findings for publication in a peer-reviewed journal in a way that is consistent with global standards.
- ✓ Guide you in preparing your study design and related monitoring and evaluation (M&E) activities with an understanding of what will be needed for transparent and complete reporting.
- ✓ Encourage and empower you to contribute to global efforts to compare findings and synthesize evidence on the impact of digital health strategies by adhering to standardized reporting guidelines.

**W**hile the ultimate goal of M&E is to both optimize implementation and generate evidence about a particular programme, disseminating the findings is also critical, as this will contribute to better understanding about the impact of digital health interventions and encourage further support and investments in digital health. To assess the relevance of research findings, and potentially replicate effective digital health interventions, readers need clear and comprehensive information that includes both the evaluation methodology and the way that the specific technologies and platforms are employed to address health needs.

The mHealth Evidence Reporting and Assessment (mERA) checklist (1) was created after it was recognized that there was a lack of adequate, systematic and useful reporting for digital health interventions and associated research studies. The tool was developed to promote clarity and completeness in reporting of research involving the use of mobile tools in health care, irrespective of the format or channel of such reporting. mERA aims to provide guidance for complete and transparent reporting on studies evaluating and reporting on the feasibility and effectiveness of digital health interventions. The checklist does not aim to support the design or implementation of such studies, or to evaluate the quality of the research methods employed. Rather, it is intended to improve transparency in reporting, promote a critical assessment of digital health research evidence, and serve as a tool to improve the rigour of future reporting of research findings.

mERA was developed as a checklist of items that could be applied by two types of audiences:

- Authors developing manuscripts that aim to report on the effectiveness of digital health interventions; and
- Peer-reviewers and journal editors reviewing such evidence.

The checklist should also be reviewed before starting any M&E activity that is planned for publication, to ensure that the necessary reporting criteria have been accounted for.

## Part 6a: How to use mERA

mERA is a checklist consisting of 16 items focused on reporting on digital health interventions; these items are listed in Table 6.1 and described in detail later in this section. As far as possible, the 16 core mERA items should be used in conjunction with appropriate study-design-specific checklists, such as the CONSolidated Standards Of Reporting Trials (CONSORT) checklist of recommendations for reporting on randomized controlled trials (2), and the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) checklists for reporting on observational studies, including cohort, case–control and cross-sectional studies (3).

In addition to the 16 core mERA criteria, the mERA authors also present a checklist of 29 items for reporting on study design and methods – these are provided in Part 6b of this Guide. These general methodology criteria were developed based on existing checklists, to specifically guide methodological reporting of digital health evidence, which has to date largely employed more exploratory study designs. We present this checklist of methodological criteria, including examples (see Table 6.2), as guidance for authors who may be unfamiliar with study-design-specific checklists, to raise awareness of important aspects of study design and implementation that should be reported, at a minimum, to allow research to undergo synthesis and meta-analysis. We reiterate here, however, the importance of also following published and accepted global guidelines for the reporting of research, in accordance with the study design and methods used.

The 16 mERA core items are presented in Table 6.1 along with the description of each item. After the table, an example of good reporting and further explanation of each item is presented.

**Table 6.1. Core mERA criteria**

Item No.	Criterion	Description
1	Infrastructure (population-level)	Presents the availability of infrastructure to support technology operations in the study location. This refers to physical infrastructure such as electricity, access to power and connectivity in the local context. Reporting X% network coverage rate in the country is insufficient if the study is not being conducted at the country level.
2	Technology platform	Describes and provides justification for the technology architecture. This includes a description of software and hardware and details of any modifications made to publicly available software.
3	Interoperability/health information systems (HIS) context	Describes how the digital health strategy can integrate into existing HIS. Refers to whether the potential of technical and structural integration into existing HIS or programme has been described irrespective of whether such integration has been achieved by the existing system.
4	Intervention delivery	Describes the delivery of the digital health intervention. This should include frequency of mobile communication, mode of delivery of intervention (i.e. SMS, face-to-face, interactive voice response), timing and duration over which delivery occurred.
5	Intervention content	Describes details of the content of the intervention. Source and any modifications of the intervention content is described.
6	Usability/content testing	Describes formative research and/or content and/or usability testing with target group(s) clearly identified, as appropriate.
7	User feedback	Describes user feedback about the intervention or user satisfaction with the intervention. User feedback could include user opinions about content or user interface, their perceptions about usability, access, connectivity.
8	Access of individual participants	Mentions barriers or facilitators to the adoption of the intervention among study participants. Relates to individual-level structural, economic and social barriers or facilitators to access such as affordability, and other factors that may limit a user's ability to adopt the intervention.
9	Cost assessment	Presents basic costs assessment of the digital health intervention from various perspectives. This criterion broadly refers to the reporting of some cost considerations for the digital health intervention in lieu of a full economic analysis. If a formal economic evaluation has been undertaken, it should be mentioned with appropriate references. A separate reporting criterion is available to guide economic reporting.
10	Adoption inputs/Programme entry	Describes how people are informed about the programme including training, if relevant. Includes description of promotional activities and/or training required to implement the digital health intervention among the user population of interest.

11	Limitations for delivery at scale	Clearly presents the digital health intervention's limitations for delivery at scale.
12	Contextual adaptability	Describes the adaptation, or not, of the solution to a different language, different population or context. Any tailoring or modification of the intervention that resulted from pilot testing/ usability assessment is described.
13	Replicability	Details the intervention to support replicability. Clearly presents the source code/ screenshots/flowcharts of the algorithms or examples of messages to support replicability of the digital health intervention in another setting.
14	Data security	Describes the data security procedures/confidentiality protocols.
15	Compliance with national guidelines or regulatory statutes	Details the mechanism used to assure that content or other guidance/information provided by the intervention is in alignment with existing national/regulatory guidelines is described.
16	Fidelity of the intervention	Describes the strategies employed to assess the fidelity of the intervention (i.e. was the intervention delivered as planned?). This may include assessment of participant engagement, use of back-end data to track message delivery and other technological challenges in the delivery of the intervention.

Source: adapted from Agarwal et al., 2016 (1).

**Item 1. Infrastructure: Describe, in detail, the necessary infrastructure which was required to enable the operation of the digital health intervention**

**Example:** "The rapid increase of teledensity, from under 3% in 2002 to 33.5% in 2010, combined with a total adult literacy rate of 75% (2008), allowed this mHealth intervention to reach a large population" (4).

**Explanation:** Have the authors clearly described the infrastructure required to support technology operations in the study location? This refers to physical infrastructure, such as electricity, access to power, connectivity in the local context. Simply reporting X% network coverage rate at a national level is likely inadequate if the study is not being conducted at the country level, given that there is usually regional variability; the national coverage rate may, however, be the only data available. Reporting the minimum infrastructural support requirements is helpful to those who wish to understand the feasibility, generalizability and replicability of the digital health innovation in other contexts, where infrastructure might be inferior to the location where the reported programme was conducted. As much as possible, authors should strive to describe the minimum enabling infrastructure required for programme implementation.

**Item 2. Technology platform: Describe, in sufficient detail to allow replication of the work, the software and hardware combinations used in the programme implementation**

**Example:** "RapidSMS® is an open source SMS application platform written in Python and Django. The SMS-based project was developed to track the pregnancy lifecycle . . . alerting health facilities, hospital and ambulances" (5).

**Explanation:** Have the authors explained the choices of software and hardware used in the deployment of the described digital health intervention? Clear communication of the technology used in the programme is critical to allow the contextualization of the authors' work as compared to other innovations. Without this information, it is difficult to group projects which have taken identical (or similar) approaches to resolving health system constraints. If the software used is a publicly available system (e.g. Open Data Kit [ODK], CommCare) it should be explicitly mentioned, together with the modifications or configuration details, with links provided to the code, if publicly available. If the application or system has been custom-coded for the programme and is open-source, the link to the public repository where the code is housed would be useful to researchers attempting to replicate the authors' work. Similarly, the hardware choices made should be described with detail akin to that in item 1 (Infrastructure), to allow implementers of future programmes to understand the minimum technical functionality required for the software performance of replicate deployments to be similar in nature to the programme being reported. For example, details on modifications such as whether the devices were functionally "locked-down" to limit use of non-study applications should be reported.



---

**Item 3. Interoperability: Describe how, if at all, the digital health intervention connects to and interacts with national or regional health information systems (HIS)/programme context**

**Example:** “Text messages were sent using a customized text-messaging platform integrated with the institution’s immunization information system” (6).

**Explanation:** Clarity of the “fit” within the existing HIS, either the national HIS or that of the host organization, is important to understanding how the digital health intervention adds to the existing workflows, improves on existing processes, or complements existing programmes. Many digital health projects have been criticized for operating in silos, independent of existing efforts to create organizational or national HIS architectures or to integrate with existing health promotion programmes (7). Simple descriptions of specific data standards being used (e.g. HL7, OpenMRS CIEL, ICD-9/10) can provide some basis for gauging a programme’s interoperability readiness, helping also to understand whether the activity is a limited-scale pilot project or an intervention being built for eventual national scale-up. The degree to which a programme may already be integrated into a national HIS may also be reported, explaining how data elements contribute to aggregate reporting through systems such as District Health Information Systems (DHIS).

---

**Item 4. Intervention delivery: Elaborate the mode, frequency and intensity of the digital health intervention**

**Example:** “Parents of children and adolescents in the intervention group received a series of 5 weekly, automated text message influenza vaccine reminders” (6).

**Explanation:** Often, in reporting the digital health intervention, authors omit important details around the specific “intervention” to which participants are being exposed. Firstly, the channels used to provide information or engage with the client should be described (e.g. SMS, voice message, USSD), as this choice may explain operational variability across similar deployments. Parameters such as the intensity and frequency of interactions, the duration of engagement, and the time of day (if relevant) should be described. For example, with a text message intervention to stimulate behaviour change, how was the message curriculum structured, timed and delivered? Was attention paid to the time of day? Were limits placed on the number of messages sent in a given week, to address concerns about information saturation? Were choices between modes of delivery offered to clients (e.g. interactive voice response [IVR] instead of text messages)? For what total duration was the intervention implemented?

---

**Item 5. Intervention content: Describe how the content was developed/identified and customized**

**Example:** “A topic message on Monday, Wednesday, Thursday, and Saturday, such as ‘Control your portions by setting aside a large snack package into smaller bags or buy 100-calorie snack packs!’” (8).

**Explanation:** We recommend that the source of any informational content (e.g. behaviour recommendations, decision-support guidelines, medication or referral recommendations, global or national technical guidelines) be mentioned clearly, together with any specific adaptation that may have been made to the content based on the location of the particular project. If new content was created, the process of enlisting qualified experts and the development, validation and testing of novel content should be described. If information content is drawn from a publicly available resource, or newly developed content is being made publicly available, external links to this database should be provided.

---

**Item 6. Usability testing: Describe how the end-users of the system engaged in the development of the intervention**

**Example:** “Designing the system began with formative research with overweight men and women to solicit feedback about dietary behaviours, current mobile phone and text and picture message habits, the type and frequency of text and picture messages helpful for weight loss, and nutrition-related topic areas that should be included in a weight loss programme” (8).

**Explanation:** Given the space limitations in most peer-reviewed journals, this important element of a carefully developed digital health innovation is often given short shrift. Separate manuscripts or documents may exist, however, describing the formative research undertaken to capture user needs, define system constraints, map user workflows, and to adapt communication content and technical solutions to fit the local context. If this is the case, the publication should provide clear reference to where such detail may be found, since this information is extremely useful to many readers attempting to either contextualize or replicate the work. The definition and recruitment of “end-users” should be clearly explained, together with a brief overview of the depth and breadth of formative work undertaken to engage them in the development of the system. Conversely, if end-users were not involved, this, too, should be explicitly mentioned.

---

**Item 7. User feedback: Describe user feedback about the intervention or user satisfaction with the intervention**

**Example:** “Most telephone respondents reported that the platform was easy to use and simple, and appreciated the ability to obtain health information via mobile phone” (9).

**Explanation:** Has user response to the digital health intervention been assessed, and acceptance verified? This information is key for documenting the likelihood of adoption of the intervention among end-users. Digital health interventions are sometimes developed without sufficient audience or end-user feedback, although their opinions are critical for assessing the likelihood of success of the digital health intervention at both the pilot and large-scale level of implementation. User feedback could include user opinions about the content or user interface, their perceptions about usability, access and connectivity, or other elements of the digital health system. User feedback should inform the reader’s understanding of how and why the digital health intervention is expected to succeed, as well as challenges that may be encountered during programme implementation and replication. Without reporting this information, a seemingly elegant digital health system may have limited likelihood of user adoption.

---

**Item 8. Access of individual participants: Mention barriers or facilitators to the adoption of the intervention among study participants**

**Example:** “It is possible that this intervention is less effective among certain subpopulations that may be considered harder to reach (i.e. males, those with a lower level of education and those who do not regularly attend health services)” (10).

**Explanation:** Have the authors considered who the digital health intervention will work for and who will face challenges in accessing it? Given the requirements for access to and use of digital health systems, some population subgroups may be more or less likely to adopt it. As with all modes of delivering health interventions, limitations of access among certain subgroups are likely and therefore should be candidly considered in the peer-reviewed report. Challenges to access may relate to socioeconomic status, geographic location, education and literacy, gender norms that limit access to resources and information, as well as a host of other demographic and sociocultural factors. Discussion of potential limitations to access will help the reader to make an informed assessment of whether the digital health intervention is appropriate for other target groups.



---

**Item 9. Cost assessment: Present basic costs of the digital health intervention**

**Example:** “Health workers in Salima and Nkhotakota with no access to the SMS program tend to spend an average of 1,445 minutes (24 hours) to report and receive feedback on issues raised to their supervisor at an average cost of USD \$2.70 (K405.16) per contact, and an average contact frequency of 4 times per month” (11).

**Explanation:** Recognizing the rarity of full economic evaluations, we propose that basic information on financial costs required to design/develop, start-up and sustain implementation should be reported, from the perspective of different users of the system, over a specified time period. Ideally these perspectives would include programme, health systems, mobile network operator and end-user costs. Rarely do digital health programme reports provide such cost information, essential to programme evaluation and selection. Full economic evaluations entail the comparison of costs and consequences for two or more alternatives and may include cost–effectiveness, cost–utility, cost–benefit, or cost–minimization analyses. If a full economic evaluation has been conducted, it should be reported according to the 24-item Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement (12).

---

**Item 10. Adoption inputs/ programme entry: Describe how people are informed about the programme or steps taken to support adoption**

**Example:** “Training on how to use the cell phones and on text-messaging protocol took place in 2 2-hour sessions on consecutive days. The first day involved training on how to use the cell phone – using pictographic instructions and interactive exercises – which was conducted in small groups (3–6 participants) and facilitated by a bilingual (English and Twi) proctor” (13).

**Explanation:** Appropriate training, instructional materials and competency assessment may be warranted as digital health strategies typically require the health-care provider or client end-users to have a level of understanding of the scenarios of use and the relevant available functionality, as well as the competence to be able to appropriately use the intervention. Have the authors provided a description of the instructional approaches deployed for end-users of the digital health system, or justification for their exclusion? Authors should ensure that the details of these inputs are described, including: for health workers, validity of instructional approach utilized, competency of instructors, validation of instructional materials, numbers of participants per session, number and length of instruction, and use of user-guides and competency assessment tools; for clients, instructional user-guide materials and/or training, length and frequency of training, and use of competency assessment tools. If instructional materials are available publicly, access details should be provided.

---

**Item 11. Limitations for delivery at scale: Present expected challenges for scaling up the intervention**

**Example:** “Despite our findings that the intervention was not burdensome and was indeed well-accepted by health workers, sending 2 messages daily for 5 days a week over 26 weeks to each health worker leaves limited space for other similar, non-malaria quality improvement interventions” (14).

**Explanation:** Given the challenges in translating findings from pilot studies to large-scale implementations, authors should recognize the limiting factors surrounding delivery at scale. Pilot studies often maintain implementation fidelity and activities are closely monitored at a level that may not be sustained during a large-scale implementation. Have the authors discussed the level of effort involved in the implementation and considered the constraints relevant to further scale-up of the intervention? This information is critical for understanding the generalizability of the implementation and making inferences on its viability beyond a closely controlled and defined setting.

**Item 12. Contextual adaptability: Describe appropriateness of the intervention to the context, and any possible adaptations**

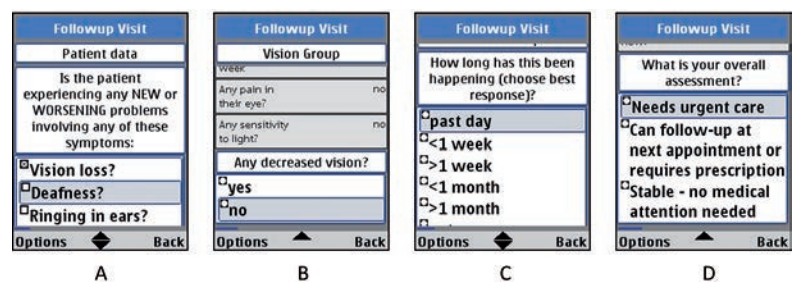
**Example:** “Our mobile phone based survey apparatus may be particularly suited for conducting survey research in rural areas. In surveys where multiple research sites may be remote and dispersed, and where vehicles have to be used to travel from site to site to download data onto laptops, the mobile phone based data collection system may be a significantly cheaper option” (15).

**Explanation:** The digital health intervention may have functionality that is broadly applicable to a range of settings and usage scenarios, and specific functionality that is only suited to specific needs, users and geographic localities. Have the authors provided details of the relevance of the functionality of the digital health intervention to the specific research context, and drawn inferences of potential relevance and adaptability based on health domains, user types, geographic contexts and health needs? Have the authors described the steps necessary to adapt the digital health intervention to other use cases? In some cases, if a piece of software is hard-coded, adaptability may be limited, costly or time-consuming. Specifying limitations to the contextual adaptability of the system being reported helps to clarify whether the system being tested can be considered a potential “platform”, useful to multiple future purposes, or if the system was designed specifically as a single-use proof-of-concept.

**Item 13. Replicability: Present adequate technical and content detail to support replicability**

**Example:** “The mobile phone application, CommCare, developed by Dimagi, Inc., was iteratively modified into Mobilize” (16).

**Figure 6.1: Screen shot images of Mobilize on the mobile phones**



**Explanation:** The potential for a digital health intervention to be efficiently introduced to a new population is enhanced by the development and availability of standard operating procedures of successful interventions. Have the authors provided details of the development of replicable processes that are being deployed in a consistent manner? These may include the software source code, screenshots of the workflow/dashboards, flowcharts of algorithms, or examples of content that has been developed for the end-users. If this level of detail cannot be included in the manuscript due to space restrictions, links to external resources should be provided.

---

**Item 14. Data security: Describe security and confidentiality protocols**

**Example:** “All survey data were encrypted, thus maintaining the confidentiality of responses. Communication between the browser and the server was encrypted using 128-bit SSL. System servers were secured by firewalls to prevent unauthorized access and denial of service attacks, while data was protected from virus threats using NOD32 anti-virus technology” (15).

**Explanation:** A brief explanation of the hardware, software and procedural steps taken to minimize the risk of data loss or data capture should be reported. Many ethical review bodies are now requiring investigators to report the details of steps taken to secure personal information that can be linked to a client’s identity, from identity fields to laboratory test results. Even in settings where laws, standards or practices governing data security may be absent, researchers and programme implementers are responsible for taking reasonable measures to protect the privacy and confidentiality of participants’ identity and health information. Data security reporting should cover measures taken at the time of collection/capture of information, transmittal of information, receipt and storage of information, as well as any access control measures that are in place. Data-sharing protocols, if any, should be mentioned in this section.

---

**Item 15. Compliance with national guidelines or regulatory statutes: Details mechanism used to assure that content provided by the intervention is in alignment**

**Example:** “The research assistant programmed the message into the automated, web-based, and HIPAA<sup>8</sup> compliant Intelecare platform” (17).

**Explanation:** If the digital health intervention or application is being used to deliver health information or decision-support guidance, or to provide diagnostic support to health workers, the authors should describe whether national guidelines or another authoritative source of information has been used to populate system content. For example, if the system is providing SMS-based advice to pregnant women, does the information follow evidence-informed practices and align with recommendations from existing national/regulatory bodies? In some jurisdictions, the provision of health-care advice or treatment guidelines falls under the specific oversight of a national agency, such as the Federal Communications Commission (FCC) or Food and Drug Administration (FDA) in the United States. This is especially true when the technology can be considered a “medical device”. If this determination has been made, and if specific regulatory oversight has been sought, this should be reported.

---

**Item 16. Fidelity of the intervention**

**Example:** “On average, users transferred data manually (pressed the button) 0.9 times a day, where the most eager user transferred data 3.6 times a day and the least eager none. Six of the 12 users experienced malfunctions with the step counter during the test period – usually a lack of battery capacity or an internal ‘hang-up’ in the device that needed a hard restart” (18).

**Explanation:** To what extent has the digital health programme’s adherence to the intended, original deployment plan been assessed? If systems have been put in place to monitor system stability, ensure delivery (and possibly, receipt) of messages, or measure levels of participant/end-user engagement with the system, these can generate metrics of intervention fidelity. Gaps in fidelity assessment and reporting make it difficult to link intervention delivery to possible process or health outcomes. Fidelity metrics may be based on either system-generated data, monitoring data or a combination of both.

---

8 The Health Insurance Portability and Accountability Act (HIPAA) is set of regulations governing data privacy and security provisions for safeguarding medical information. Further information is available at: <http://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>

## Part 6b: Methodological criteria

The checklist in Table 6.2 is sub-divided into two areas: (a) essential criteria for all studies; (b) essential criteria by type of study, the latter being differentiated by qualitative and quantitative research methods. This list of methodological criteria was based on a comprehensive search for published guidelines for reporting quantitative and qualitative studies. Based on the study design, one or more of these sub-domains can be applied. If the study has used a mixed-methods approach, comprising both qualitative and quantitative research techniques, both sets of criteria listed in section B of Table 6.2 should be applied.

**Table 6.2. mERA Methodological criteria**

Criteria	Item No.	Description	Examples
<b>A. ESSENTIAL CRITERIA FOR ALL STUDIES</b>			
<b>Introduction</b>			
<b>Rationale/scientific background</b>	1	Mentions previous studies or scientific documentation (relating to a similar or different context) that have attempted to answer the question. States the rationale for the study.	"Randomized controlled trials of such use of text messaging are scarce, with only two trials from low resource settings" (19).
<b>Objectives/hypotheses</b>	2	States specific objectives of the study, including whether a hypothesis is being tested.	"This study describes findings from formative research to examine pre-intervention community and health worker perceptions on use of mobile phones to improve PMTCT-related communication" (20).
<b>Logic model/theoretical framework</b>	3	Depicts a theoretical framework/logic model on how the intervention influences the primary study outcome. If the intervention is targeting behaviour change, appropriate behaviour change theory is described.	"In order to understand the factors influencing the adoption of technology in healthcare, we utilized the ICT for healthcare development (ICT4H) model to illustrate the key benefits, and related barriers, of mobile phones as a healthcare tool" (21).
<b>Methods</b>			
<b>Study design</b>	4	Presents a description of the study design and how it was arrived at.	"The study is a pragmatic cluster-randomized controlled trial with the primary healthcare facility as the unit of randomization" (22).
<b>Outcomes</b>	5	Defines primary and secondary outcome measures clearly related to study objectives. Mentions secondary outcome measures, if relevant.	"The primary outcome was correct management with artemether-lumefantrine, defined as a dichotomous composite indicator of treatment, dispensing, and counselling tasks concordant with Kenyan national guidelines" (19).
<b>Data collection methods</b>	6	Provides a description of data collection methods; this could include description of the study tools/survey questionnaires/interview guides.	"This 37 item questionnaire was modified from the Eating Behavior Inventory (EBI) by adding 10 items. While the EBI has good reliability and validity, additional questions assessed weight related behaviors (e.g. goal setting, food choices, barriers) not captured by the EBI" (23).
<b>Participant eligibility</b>	7	Describes eligibility criteria for participants.	"Men eligible for enrolment were 18 years of age or older, had undergone circumcision on the day of screening, owned a mobile phone, had the phone in their possession at the time of enrolment, and were able and willing to respond to a questionnaire administered by phone 42 days after circumcision" (24).
<b>Participant recruitment</b>	8	Provides a description of how the study participants were recruited into the study; this might include self-selection, health-care facility-based recruitment or community-based recruitment, among others.	"Between September 2010 and April 2011, 1,200 men were randomly assigned to receive either the intervention or standard care. Men who had undergone circumcision were approached by study staff during the 30-minute post-operative recovery period" (24).

Criteria	Item No.	Description	Examples
<b>Bias</b>	9	Reports risk of bias. Key examples of bias include recall bias (error in the accuracy or completeness of respondent recollections) and selection bias (errors in selection of participating individuals or groups).	"Although the reported outcomes may be accurate and reflect the proximity of the targeted communities to the Sene District Medical Center, they may also reflect underreporting biases of birth attendants who may have omitted reports of patients when outcomes were less favorable" (13).
<b>Sampling</b>	10	Explains how the sample size was determined and whether attrition was accounted for, if relevant to the study design.	"To provide at least 90% power to detect a relative risk (RR) between the two arms of 1.22 or larger for return to clinic (equivalent to an increase from 43% to 52.5%), we needed to enroll 1200 men (600 in each arm)" (24).
<b>Setting and locations</b>	11	Provides a description of the study population, as well as details on geographic area and context. Please note that data could be collected in a sub-sample of a larger geographic area/ population where the intervention is being implemented. This concept thus refers to details on the population/context from which data were collected.	"The study took place in 2009–10 on the island of Unguja in Zanzibar, a part of the United Republic of Tanzania. The island has six districts with 80 healthcare facilities, 95% of which are government owned. Of the six districts, two are urban (Urban and West) and four are rural (North A, North B, Central, South)" (22).
<b>Comparator</b>	12	Describes use of a comparison group. For studies that do not use a comparator, this would not be relevant. Please note that before/after comparisons are also valid and should be noted.	"To ensure comparability of the intervention and women with respect to socioeconomic baseline characteristics, two of the four selected facilities in each district were randomly assigned to intervention (mobile phone intervention) and two to control (standard care)" (22).
<b>Data sources/ measurement</b>	13	Describes the source of data for each variable of interest. Data source may include individuals, databases, search engines.	"Two main data sources were used to evaluate feasibility, reach and potential impact on family planning use. First, to evaluate feasibility, all queries to the m4RH system during the pilot period were tracked through electronic and automatic logging of each system query" (25).
<b>Results</b>			
<b>Enrolment</b>	14	Describes enrolment procedures. Mentions the numbers of participants screened for eligibility, found to be eligible/not eligible, declined to be enrolled, and enrolled in the study, as relevant.	"Table 6.1 shows the number of children enrolled into the study at each survey, and the number of children included in intention-to-treat and per-protocol analyses. No carer of a sick child refused to participate in the study" (19).
<b>Description of study population</b>	15	Provides demographic and/or clinical characteristics of participants in each study cohort.	"The majority of the women participating were housewives and farmers of rural residence, and 17% were totally illiterate. Thirty-seven per cent of the women included in the study owned a mobile phone" (22).
<b>Reporting on outcomes</b>	16	Presents each primary and secondary outcome for study findings.	"Text-messaging interventions yielded significantly higher adherence than control conditions (OR = 1.39; 95% CI = 1.18, 1.64)" (26).
<b>Discussion</b>			
<b>Summary of evidence</b>	17	Provides general interpretation of the results in the context of current evidence and current theory.	"This is a randomized study that evaluates the effect of text message reminders on OCP continuation. The 6-month continuation rate in the control group was similar to other recent reports. Six-month OCP continuation improved by 10% in the intervention group" (27).



Criteria	Item No.	Description	Examples
<b>Limitations</b>	18	Discusses study limitations; this should address sources of potential biases and imprecision.	"Another limitation in the design was the before-after nature of the assessments as such designs have the potential for temporal confounding" (28).
<b>Generalizability/external validity</b>	19	Discusses applicability of study findings to other settings. Examples might include a discussion on study population, characteristics of the intervention, incentives, and compliance rates in other contexts.	"Although the generalizability of results from this evaluation are limited, the 40% response rate to text questions is similar to other technology-based data collection approaches, and the text data have face validity in that the type of contraceptive methods queried and reported aligned well with the age of users" (25).
<b>Conclusions/interpretation</b>	20	Provides interpretation of the results. Discussion of barriers and/or opportunities relating to policy, programmes or research.	"Based on this finding, the authors anticipate a much broader use of mobile technology for the delivery of clinical standards aimed at improving clinical care in low-income countries, including studies investigating the benefits with other clinical protocols and with other cadres of healthcare workers" (28).

#### Conflicts

<b>Funding</b>	21	Lists sources of funding and role of funders.	"This study and the development of this article was supported in part by research grant 1 K01 TW008763-01A1 from Fogarty International, National Institutes of Health" (29).
<b>Ethical considerations</b>	22	Addresses the process of reviewing the ethical issues related to participant enrolment, including obtaining consent and preservation of confidentiality.	"This study was approved by the Kenya Medical Research Institute's Ethical Review Committee, the University of Washington's Human Subjects Division, and Institutional Review Board #3 at the University of Illinois at Chicago" (24).
<b>Competing interests</b>	23	Describes any conflicts of interest.	

#### B. ESSENTIAL CRITERIA BY TYPE OF STUDY

##### B.1 Quantitative study

<b>Confounding</b>	1	Reports the risk of confounding and any methods used to address this.	"Any factor that changed either of the two difference-of-differences effect sizes by more than 20% was regarded as a confounder and retained in the final model" (19).
<b>Statistical methods</b>	2	Describes methods for primary and additional analyses, such as subgroup analyses and adjusted analyses.	"Using an intent-to-treat approach, we tested the hypothesis that educational daily text messages would affect oral contraceptive continuation. We used Student's <i>t</i> test to compare continuous variables and Pearson's $\chi^2$ test to compare categorical variables when describing the population and assessing characteristics associated with 6-month OCP continuation" (27).
<b>Missing data</b>	3	Reports methods for dealing with missing data due to incomplete surveys; this refers to how the data for missing, non-response and other variables were handled.	"We used multiple imputation (PROC MI and MIANALYZE in SAS, with five sets of imputations) to impute the missing values of covariates" (30).

##### B.2 Qualitative criteria

<b>Analytical methods</b>	1	Describes analytical methods, including in-depth description of the analysis process and how categories/themes were derived.	"In the second phase, a manual preliminary analysis of the narrative data aimed to assemble the responses according to the pre-set themes in the FGD topic guide, which were then refined according to emergent themes" (20).
---------------------------	---	--	---



Criteria	Item No.	Description	Examples
<b>Data validation</b>	2	Discusses use of triangulation, member checking (respondent validation), search for negative cases, or other procedures for cross-verification from two or more sources.	"To enhance the credibility of results, the research team compared findings from each of the study's sub-groups and organized a verification meeting with program advisors to examine the extent to which the research captured internally valid and dependable information" (20).
<b>Reflexivity of account provided</b>	3	Describes the researcher's role and relationship to the respondent, wording and phrasing of questions and other factors that might have elicited a biased response.	"The three authors, all women ranging in age from mid-20s to late-50s, can be considered to fall on a continuum between "outsider" and "insider". The first author moved to El Paso to accept a university position and is not of Mexican origin; however, she has been involved with Gente a favor de gente for several years, along with the second author" (31).

Source: The first three columns of this table were adapted from Web appendix 1: mERA methodology criteria, in: Agarwal et al., 2016 (1).

This list of methodological criteria was based on six sources that were selected after a comprehensive search for published guidelines for reporting quantitative and qualitative studies (32–37).

## References

1. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ*. 2016;352:i1174. doi:10.1136/bmj.i1174.
2. CONSORT checklist. In: CONSolidated Standards Of Reporting Trials (CONSORT) [website]. CONSORT; 2016 (<http://www.consort-statement.org/checklists/view/32-consort/66-title>, accessed 15 April 2016).
3. STROBE checklists. In: STROBE STrengthening the Reporting of OBservational studies in Epidemiology [website]. STROBE; 2016 (<http://www.strobe-statement.org/>, accessed 15 April 2016).
4. Jamison JC, Karlan D, Raffler P. Mixed-method evaluation of a passive mHealth sexual information texting service in Uganda. *Inf Technol Int Dev*. 2013;9(3):1–28 (<http://itidjournal.org/index.php/itid/article/view/1087>, accessed 16 May 2016).
5. Ngabo F, Nguimfack J, Nwaigwe F, Mugeni C, Muhoza D, Wilson DR et al. Designing and implementing an innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. *Pan Afr Med J*. 2012;13:31.
6. Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. *JAMA*. 2012;307:1702–8.
7. The PLoS Medicine Editors. A reality checkpoint for mobile health: three challenges to overcome. *PLoS Med*. Public Library of Science; 2013;10(2):e1001395. doi:10.1371/journal.pmed.1001395.
8. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL et al. A text message-based intervention for weight loss: randomized controlled trial. *J Med Internet Res*. 2009;11:e1.
9. Vahdat HL, L'Engle KL, Plourde KF, Magaria L, Olawo A. There are some questions you may not ask in a clinic: providing contraception information to young people in Kenya using SMS. *Int J Gynecol Obstet*. 2013;123:e2–6.
10. Gold J, Lim M, Hocking J, Keogh LA, Spelman T, Hellard ME. Determining the impact of text messaging for sexual health promotion to young people. *Sex Transm Dis*. 2011;38:247–52. doi:10.1097/OLQ.0b013e3181f68d7b.
11. Lemay NV, Sullivan T, Jumbe B, Perry CP. Reaching remote health workers in Malawi: baseline assessment of a pilot mHealth intervention. *J Health Commun*. 2012;17:105–17. doi:10.1080/10810730.2011.649106.

12. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMC Med*. 2013;11(1):80. doi:10.1186/1741-7015/11/80.
13. Andreatta P, Debpuur D, Danquah A, Perosky J. Using cell phones to collect postpartum hemorrhage outcome data in rural Ghana. *Int J Gynecol Obstet*. 2011;113:148–51. doi:10.1016/j.ijgo.2010.11.020.
14. Zurovac D, Larson B, Sudoi RK, Snow RW. Costs and cost-effectiveness of a mobile phone text-message reminder programmes to improve health workers' adherence to malaria guidelines in Kenya. *PLoS One*. 2012;7(12):e52045. doi:10.1371/journal.pone.0052045.
15. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. *BMC Med Inform Decis Mak*. 2009;9:51. doi:10.1186/1472-6947-9-51.
16. Chaityachati KH, Loveday M, Lorenz S, Lesh N, Larkan L-M, Cinti S et al. A pilot study of an mHealth application for healthcare workers: poor uptake despite high reported acceptability at a rural South African community-based MDR-TB treatment program. *PLoS One*. 2013;8(5):e64662. doi:10.1371/journal.pone.0064662.
17. Dowshen N, Kuhns LM, Gray C, Lee S, Garofalo R. Feasibility of interactive text message response (ITR) as a novel, real-time measure of adherence to antiretroviral therapy for HIV+ youth. *AIDS Behav*. 2013;17(6):2237–43.
18. Årsand E, Tataru N, Østengen G, Hartvigsen G. Mobile phone-based self-management tools for type 2 diabetes: the few touch application. *J Diabetes Sci Technol*. 2010;4(2):328–36.
19. Zurovac D, Sudoi RK, Akhwale WS, Ndiritu M, Hamer DH, Rowe AK et al. The effect of mobile phone text-message reminders on Kenyan health workers' adherence to malaria treatment guidelines: a cluster randomised trial. *Lancet*. 2011;378(9793):795–803.
20. Jennings L, Ong J, Simiyu R, Sirengo M, Kassaye S. Exploring the use of mobile phone technology for the enhancement of the prevention of mother-to-child transmission of HIV program in Nyanza, Kenya: a qualitative study. *BMC Public Health*. 2013;13(1):1131.
21. Chib A, Jia CY, Lee Lin Chieh L, Ng Chiah Hwee C, Kee TANC, Kameswari VLV. The hope of mobile phones in Indian rural healthcare. *J Health Inform Dev Ctries*. 2012;6(1).
22. Lund S, Hemed M, Nielsen BB, Said A, Said K, Makungu MH et al. Mobile phones as a health communication tool to improve skilled attendance at delivery in Zanzibar: a cluster-randomised controlled trial. *BJOG An Int J Obstet Gynaecol*. 2012;119(10):1256–64.
23. Shapiro JR, Koro T, Doran N, Thompson S, Sallis JF, Calfas K et al. Text4Diet: a randomized controlled study using text messaging for weight loss behaviors. *Prev Med*. 2012;55(5):412–7.
24. Odeny TA, Bailey RC, Bukusi EA, Simoni JM, Tapia KA, Yuhas K et al. Text messaging to improve attendance at post-operative clinic visits after adult male circumcision for HIV prevention: a randomized controlled trial. *PLoS One*. 2012;7(9):e43832.
25. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. *Contraception*. 2013;87(2):251–6. doi:10.1016/j.contraception.2012.07.009.
26. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. *PLoS One*. 2014;9(2):e88166.
27. Castaño PM, Bynum JY, Andrés R, Lara M, Westhoff C. Effect of daily text messages on oral contraceptive continuation. *Obstet Gynecol*. 2012;119(1):14–20.

28. Mitchell M, Hedt-Gauthier BL, Msellemu D, Nkaka M, Lesh N. Using electronic technology to improve clinical care – results from a before-after cluster trial to evaluate assessment and classification of sick children according to Integrated Management of Childhood Illness (IMCI) protocol in Tanzania. *BMC Med Inform Decis Mak.* 2013;13(1):1–8.
29. Lori JR, Munro ML, Boyd CJ, Andreatta P. Cell phones to collect pregnancy data from remote areas in Liberia. *J Nurs Scholarsh.* 2012;44(3):294–301.
30. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. *Lancet.* 2010;376(9755):1838–45.
31. Grineski SE, Hernández AA, Ramos V. Raising children in a violent context: an intersectionality approach to understanding parents' experiences in Ciudad Juárez. *Womens Stud Int Forum.* 2013;40:10–22.
32. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health.* 2004;94(3):361–6. doi:10.2105/AJPH.94.3.361.
33. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Prev Med.* 2007;45(4):247–51. doi:10.1016/j.ypmed.2007.08.012.
34. Meyer GJ. Guidelines for reporting information in studies of diagnostic test accuracy: The STARD Initiative. *J Pers Assess.* 2010;81(3):191–3. doi:10.1207/S15327752JPA8103.
35. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med.* 2010;8(18):32. doi:10.1186/1745-6215-11-32.
36. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care.* 2007;19(6):349–57. doi:10.1093/intqhc/mzm042.
37. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S. Publication guidelines for quality improvement in health care: evolution of the SQUIRE project. *Qual Saf Health Care.* 2008;17(Suppl 1):i3–i9. doi:10.1136/qshc.2008.029066.

# Annex I: Glossary

**Analytic study:** A study aimed at quantifying the relationship between the intervention and the outcome(s) of interest, usually with the specific aim of demonstrating a causative link between the two. These studies are designed to test hypotheses that have usually been generated from descriptive studies. There are two main categories of analytic studies: (a) experimental and (b) observational.

**Benchmark:** Reference point or standard against which performance or achievements can be assessed (1).

**Claim:** A statement of anticipated benefits of the digital health system or intervention.

**Codebook (also known as a data dictionary):** A description about a data set that details features such as the meaning, relationships to other data, origin, usage and format of specific data elements (2).

**Conceptual framework (also known as theoretical or causal framework):** A diagram that identifies and illustrates the relationships among factors (systemic, organizational, individual or other) that may influence the operation of an intervention and the successful achievement of intervention's goals (3). The purpose is to facilitate the design of the digital health intervention or project and provide a theoretical basis for the approach.

**Descriptive study:** A study that is "concerned with and designed only to describe the existing distribution of variables, without regard to causal or other hypotheses" (4).

**Digital health:** The use of digital, mobile and wireless technologies to support the achievement of health objectives. Digital health describes the general use of information and communications technologies (ICT) for health and is inclusive of both mHealth and eHealth.

**Digital health intervention:** In the context of this Guide, the application of digital, mobile and wireless technologies for a defined health purpose, in order to address specific health system challenges. For example, a digital health intervention can be text messaging to deliver messages to pregnant women for antenatal care follow-up.

**Effectiveness:** In the context of this Guide, the ability of a digital health intervention to achieve the intended results in a non-research (uncontrolled) setting.

**Efficacy:** In the context of this Guide, the ability of a digital health intervention to achieve the intended results in a research (controlled) setting.

**Evaluation:** The systematic and objective assessment of an ongoing or completed intervention with the aim of determining the fulfilment of objectives, efficiency, effectiveness, impact and sustainability (1). In this Guide (i.e. in the context of digital health interventions), evaluation is used to refer to measures taken and analysis performed in order to assess (i) the interaction of users or a health system with the digital health intervention strategy, or (ii) changes attributable to the digital health intervention.

**Experimental studies:** Studies that aim to assess the effects of a treatment or intervention that has been intentionally introduced on an outcome or outcomes of interest. Examples of experimental studies include randomized controlled trials (RCTs) and quasi-experimental studies.

**Feasibility:** The ability of a digital health system to work as intended in a given context.

**Fidelity:** A measure of whether or not an intervention is delivered as intended (5). In this Guide, fidelity is viewed from both technical and user perspectives.

**Focus group discussions (FGDs):** A type of qualitative research method used when researchers are interested in the opinions or perspectives of a group of individuals (6–12). FGDs may be used to get feedback before, during or after a project, to reach groups that are underrepresented in surveys, to compare and contrast norms between groups, and for goal-setting and prioritization (6–13).

**Formative evaluation:** Studies aimed at informing the development and design of effective intervention strategies. They may be conducted before or during implementation (14).

**Functionality (also referred to as functional suitability):** A “characteristic that represents the degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions” (15). In this Guide, functionality refers to the ability of the digital health system to support the desired intervention.

**Hierarchy of study designs:** A ranking of study designs from highest to lowest, based on their potential to eliminate bias (16).

**Human-centered design:** A process in which the needs, wants and limitations of end-users of a product are given extensive attention during its design and development (17).

**Impact:** The medium- to long-term effects produced by an intervention; these effects can be positive and negative, intended and unintended (1).

**Impact evaluations:** Studies that aim to assess the effect of the intervention on outcomes and the impact on the intended beneficiaries or clients. These evaluations require a counterfactual and draw on data generated internally (i.e. inputs, processes and outputs) as well as data on outcomes external to the project (18).

**Implementation research:** Research that “seeks to understand and work in real-world or usual practice settings, paying particular attention to the audience that will use the research, the context in which implementation occurs, and the factors that influence implementation” (19). In this Guide, implementation research refers to the assessment of the uptake, integration and sustainability of the evidence-based digital health intervention in a given context, including policies and practices.

**In-depth interviews (IDIs):** The process of eliciting detailed perspectives, opinions, experiences and feelings of individuals (6, 10, 11, 13). IDIs may be conducted over the phone or in person.

**Indicator:** A “quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor” (1).

**Inputs:** The financial, human, material or intellectual resources used to develop and implement an intervention. In this Guide, inputs encompass all resources that go into a digital health intervention.

**Logical framework (also known as logic model):** A management and measurement tool that summarizes what a project intends to do and how, what the key assumptions are, and how outputs and outcomes will be monitored and evaluated. The aim of a logic model is to clarify programme objectives and aid in the identification of expected causal links between inputs, processes, outputs, outcomes and impacts (20).

**Mobile health (mHealth):** The use of mobile and wireless technologies to support the achievement of health objectives (21).

**Monitoring:** See *Process monitoring*.

**Monitoring burden:** The amount of effort and resources required to successfully monitor the intervention; this burden is driven by the stage of maturity, the size of the implementation, the amount of data, and the number of users and indicators to be monitored.

**Non-probability sampling:** This is a sampling method that does not rely on randomization and allows higher probabilities for certain individuals to be selected for the survey.

**Observation:** The process and techniques used for observing and documenting the daily experiences, actions and situations of the population of interest in their everyday environments (10).

**Observational studies:** Non-experimental studies in which “the investigator does not intervene but rather simply ‘observes’ and assesses the strength of the relationship between an exposure and disease variable” (22).

**Outcomes:** The intermediate changes that emerge as a result of inputs and processes. Within digital health, these may be considered according to three levels: health systems, provider and client.

**Outputs:** The direct products/deliverables of process activities in an intervention (23). From a digital health perspective, outputs can include improvements in performance and user adoption.

**Participatory action research (PAR):** The use of quantitative, qualitative or mixed-methods approaches in a manner that prioritizes the role of participants in all aspects of research and implementation (24–26).

**Performance:** The degree to which an intervention or implementation team operates according to specific criteria/standards/guidelines or achieves results in accordance with stated goals or plans (1). In this Guide, performance is a measure of how well digital health users are delivering an intervention or how well the system is functioning.

**Pilot:** A small-scale deployment that allows project managers to assess implementation factors such as feasibility, acceptability and cost. In digital health, the results of a pilot study are generally used to inform decisions surrounding scaling up (21).

**Probability sampling:** A sampling method that involves a form of random selection and gives all individuals in the population an equal chance (probability) of being selected for the survey (27).

**Processes:** The activities undertaken in the delivery of an intervention – a digital health intervention for the purposes of this Guide. Processes may include training activities, partnership meetings, as well as the activities required to test and update the digital health system based on user response.

**Process monitoring:** The continuous process of collecting and analysing data to compare how well an intervention is being implemented against expected results (1). In this Guide (i.e. in the context of digital health interventions), “monitoring” and “process monitoring” are used interchangeably to refer to the routine collection, review and analysis of data, either generated by digital systems or purposively collected, which measure implementation fidelity and progress towards achieving intervention objectives.

**Quality:** A measure of the excellence, value, conformance to specifications, conformance to requirements, fitness for purpose, and ability to meet or exceed expectations (28). In this Guide, the quality of a digital health intervention is viewed from both a user and intervention content perspective.

**Quality assurance (QA) test cases:** Short sentences or paragraphs that describe expected functionality of discrete system functions and the steps to follow to perform each function. QA test cases break the more narrative or graphical functionality descriptions from the use cases and SRS into single-statement functions and expected actions.

**Randomized controlled trial (RCT):** A type of experimental study designed to assess the efficacy or effectiveness of an intervention by comparing the results in a group of subjects receiving the intervention with the results in a control group, where allocation to the intervention and control groups has been achieved by randomization.

**Results framework:** A “graphic representation of a strategy to achieve a specific objective that is grounded in cause-and-effect logic” (29). The main purpose of this type of framework is to clarify the causal relationships that connect the incremental achievement of results to intervention impact.

**Software requirements specification (SRS):** A document that outlines the technical requirements of a desired system, clearly outlining requirements from the team of health experts and project managers for the developers responsible for creating the project’s digital health system.

**Stability:** The likelihood that a technical system’s functions will not change or fail during use. In this Guide, stability refers to the ability of the digital health system to remain functional both under both normal and anticipated peak conditions for data loads.



**Stakeholders:** Entities (individuals or organizations) that have a vested interest in the digital health system or intervention, in the capacity of being a decision-maker, project staff or end-user (30).

**Summative evaluation:** A study conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced (1).

**Theory of change:** A theory of change is a causal model that links outcomes and activities to explain how and why the desired change is anticipated to occur (31). Theory-based conceptual frameworks are similar to logic models but aim to provide a greater understanding of the complex relationship between programme activities and anticipated results.

**Usability:** The “degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (32).

**Use cases:** Narrative descriptions of how a target user performs a specific task using the technology and how the system is expected to respond to each case (33).

**Users:** The individuals who directly utilize the technology using their digital devices, either to deliver health services (e.g. community health workers, district managers, clinicians) or to receive services (i.e. clients, patients).

**Value proposition:** A statement describing the benefits to end-users, with an implicit comparator, which can be a non-digital intervention or an alternative digital product (34).

**Wireframes:** Simple, schematic illustrations of the content, layout, functions and behaviour of the target system (35); they are useful in illustrating the expected functionality of a system.

## References

1. WHO evaluation practice handbook. Geneva: World Health Organization; 2013 ([http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf), accessed 5 May 2016).
2. Hagan D, Uggowitz S. Information and communication technologies for women’s and children’s health: a planning workbook for multi-stakeholder action. Geneva: The Partnership for Maternal, Newborn & Child Health (PMNCH); 2014. ([http://www.who.int/pmnch/knowledge/publications/ict\\_mhealth.pdf](http://www.who.int/pmnch/knowledge/publications/ict_mhealth.pdf), accessed 2 September 2015).
3. Monitoring and evaluation frameworks (3 parts). In: Virtual Knowledge Centre to End Violence against Women and Girls [website]. UN Women; 2012 (<http://www.endvawnow.org/en/articles/335-monitoring-and-evaluation-frameworks-3-parts.html>, accessed 3 May 2016).
4. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet*. 2002;359(9301):145-9.
5. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. *Implementation Science*. 2007;2:40.
6. Meadows KA. So you want to do research? 3. An introduction to qualitative methods. *Br J Community Nurs*. 2003;8(10):464-9.
7. Appraising scientific evidence: qualitative versus quantitative research. The Association of Faculties of Medicine of Canada; undated (<http://phprimer.afmc.ca/Part2-MethodsStudyingHealth/Chapter5AssessingEvidenceAndInformation/Appraisingscientificevidencequalitativeversusquantitativeresearch>, accessed 6 May 2016).
8. Creswell JW. *Research design: qualitative, quantitative and mixed methods approaches*, second edition. SAGE Publications; 2003.
9. Qualitative versus quantitative research. Xavier University Library; 2012 ([http://www.xavier.edu/library/students/documents/qualitative\\_quantitative.pdf](http://www.xavier.edu/library/students/documents/qualitative_quantitative.pdf), accessed 6 May 2016).

10. Mack N, Woodsong C, MacQueen KM, Guest G, Namey E. Qualitative research methods: a data collector's field guide. Research Triangle Park (NC): Family Health International; 2005 (<https://www.fhi360.org/resource/qualitative-research-methods-data-collectors-field-guide>, accessed 6 May 2016).
11. Stewart K, Gill P, Chadwick B, Treasure E. Qualitative research in dentistry. *Br Dent J*. 2008;204(5):235–9.
12. Qualitative research: getting started [website]. Durham (NC): Duke University Libraries; 2015 (<http://guides.library.duke.edu/qualitative-research>, accessed 6 May 2016).
13. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. *Br Dent J*. 2008;204(6):291–5.
14. Building our understanding: key concepts of evaluation – what is it and how do you do it? Atlanta (GA): United States Centers for Disease Control and Prevention, Health Communities Program; undated.
15. Functional suitability. In: ISO 25000 software product quality [website]. ISO 25000; 2015 (<http://iso25000.com/index.php/en/iso-25000-standards/iso-25010/58-functional-suitability>, accessed 9 May 2016).
16. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health*. 2002;56(2):119–27.
17. Glossary. In: Toolkits by K4Health [website]. Knowledge for Health (K4Health) Project; undated (<https://www.k4health.org/toolkits/mhealth-planning-guide/glossary>, accessed 6 May 2016).
18. Lance P, Guilkey D, Hattori A, Angeles G. How do we know if a program made a difference? A guide to statistical methods for program impact evaluation. Chapel Hill (NC): MEASURE Evaluation; 2014 (<http://www.cpc.unc.edu/measure/resources/publications/ms-14-87-en>, accessed 2 May 2016).
19. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. *BMJ*. 2013;347:f6753. doi:10.1136/bmj.f6753.
20. Monitoring and evaluation: some tools, methods, and approaches. Washington (DC): The International Bank for Reconstruction and Development/The World Bank; 2004.
21. The MAPS Toolkit: mHealth Assessment and Planning for Scale. Geneva: World Health Organization; 2015.
22. Song JW, Chung KC. Observational studies: cohort and case–control studies. *Plast Reconstr Surg*. 2010;126(6):2234–42.
23. Glossary of selected planning, monitoring and evaluation terms. Measurement, Learning and Evaluation (MLE) Project, Urban Reproductive Health Initiative; 2013 (<https://www.urbanreproductivehealth.org/toolkits/measuring-success/glossary-selected-planning-monitoring-and-evaluation-terms>, accessed 4 May 2016).
24. Peters D, Tran N, Adam T. Implementation research in health: a practical guide. Geneva: Alliance for Health Policy and Systems Research, World Health Organization; 2013.
25. Israel BA, Schulz AJ, Parker EA, Becker AB, Allen AJ, III, Guzman RJ. Community-based participatory research for health: from processes to outcomes (Chapter 3). In: Minkler M, Wallerstein N, editors. *Critical issues in developing and following CBPR principles*, second edition. Wiley; 2010.
26. Smith L, Rosenzweig L, Schmidt M. Best practices in the reporting of participatory action research: embracing both the forest and the trees. *Counseling Psychologist*. 2010;38(8):1115–38.
27. Probability sampling. In: Research methods knowledge base [website]. William M.K. Trochim; 2006 (<http://www.socialresearchmethods.net/kb/sampprob.php>, accessed 6 May 2016).
28. Reeves CA, Bednar DA. Defining quality: alternatives and implications. *Acad Manag Rev*. 1994;19(3):419–45 (<http://www.jstor.org/stable/258934>, accessed 11 April 2016).

29. Performance monitoring and evaluation TIPS: building a results framework. Washington (DC): United States Agency for International Development (USAID); 2010 (<https://www.ndi.org/files/Performance%20Monitoring%20and%20Evaluation%20Tips%20Building%20a%20Results%20Framework.pdf>, accessed 6 May 2016).
30. Bryson JM, Patton MQ, Bowman RA. Working with evaluation stakeholders: a rationale, step-wise approach and toolkit. *Eval Program Plann*. 2011;34(1):1–12. doi:10.1016/j.evalprogplan.2010.07.001.
31. Vogel I. Review of the use of “Theory of Change” in international development. London: United Kingdom Department of International Development; 2012 ([http://r4d.dfid.gov.uk/pdf/outputs/mis\\_spc/DFID\\_ToC\\_Review\\_VogelV7.pdf](http://r4d.dfid.gov.uk/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf), accessed 6 May 2016).
32. Usability. In: ISO 25000 software product quality [website]. ISO 25000; 2015 (<http://iso25000.com/index.php/en/iso-25000-standards/iso-25010?limit=3&start=3>, accessed 9 May 2016).
33. Use cases. In: usability.gov [website]. Washington (DC): United States Department of Health and Human Services; 2016 (<http://www.usability.gov/how-to-and-tools/methods/use-cases.html>, accessed 11 April 2016).
34. Crafting your value proposition. In: MaRS [website]. Toronto: MaRS Discovery District; 2012 (<https://www.marsdd.com/mars-library/crafting-your-value-proposition/>, accessed 21 March 2016).
35. Wireframing. In: usability.gov [website]. Washington (DC): United States Department of Health and Human Services; 2016 (<http://www.usability.gov/how-to-and-tools/methods/wireframing.html>, accessed 11 April 2016).





For more information, please contact:

**Department of Reproductive Health and Research**

World Health Organization

Avenue Appia 20, CH-1211 Geneva 27, Switzerland

Fax: +41 22 791 4171

Email: [reproductivehealth@who.int](mailto:reproductivehealth@who.int)

[www.who.int/reproductivehealth](http://www.who.int/reproductivehealth)



9 789241 511766