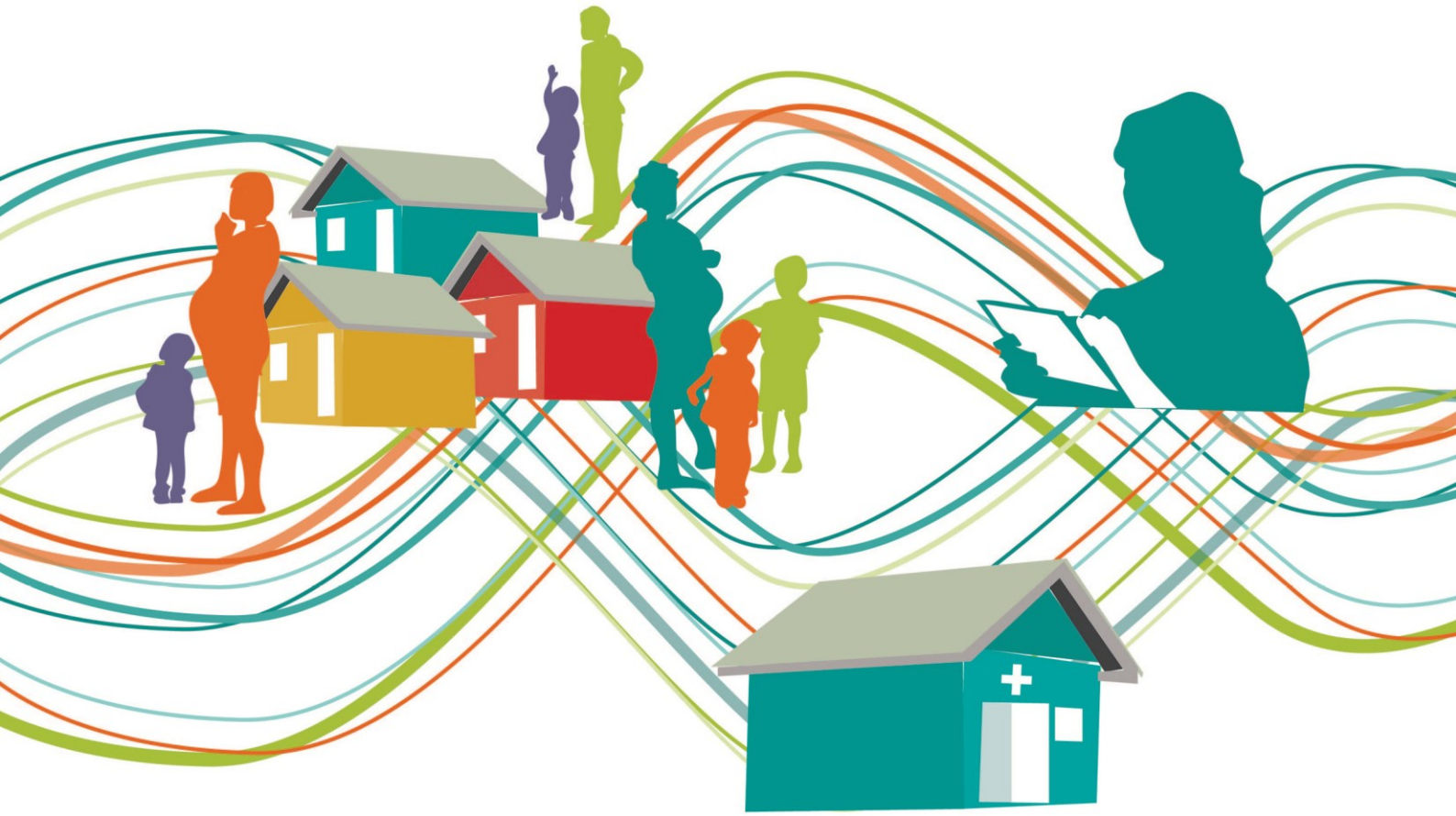


Improving Data Quality in Mobile Community-Based Health Information Systems

Guidelines for Design and Implementation

June 2017

TR-17-182

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

Improving Data Quality in Mobile Community-Based Health Information Systems

Guidelines for Design and Implementation

June 2017

MEASURE Evaluation
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, NC 27517 USA
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) under terms of Cooperative Agreement AID-OAA-L-14-00004 and implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this presentation do not necessarily reflect the views of USAID or the United States government. TR-17-182

ISBN: 978-1-9433-6459-6

USAID FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

ACKNOWLEDGMENTS

This document was prepared by Liz Nerad, Michelle Li, and Dawne Walker, all of MEASURE Evaluation/Palladium.

MEASURE Evaluation wishes to thank those who contributed to the development and publication of this assessment. First, we acknowledge the United States Agency for International Development for its support of this activity, with special thanks to Ana Djapovic Scholl. We are grateful to those who participated in stakeholder groups to share their experiences and expertise, and to those who reviewed and provided input to the final document, including colleagues from Akros, Dimagi, D-tree International, the Elizabeth Glaser Pediatric AIDS Foundation, FHI 360, ICF, IntraHealth, the Johns Hopkins University Global mHealth Initiative, John Snow Inc., Kurante, Living Goods, the Malaria Consortium, Medic Mobile, Palladium, Pathfinder, Prackelt, RTI International, the United Nations Children's Fund, and Voto Mobile. We also thank MEASURE Evaluation's knowledge management staff for their editorial support.

CONTENTS

FIGURES

TABLES

ABBREVIATIONS

ASHA	accredited social health activist
CBHIS	community-based health information system
CHW	community health worker
CTV	community trace and verify tool
HMIS	health management information system
ICCM	integrated community case management
LQAS	quality assurance sampling
mHealth	mobile health
PEPFAR	United States President's Emergency Plan for AIDS Relief
SMS	short message system
UNAIDS	Joint United Nations Programme on HIV/AIDS
VHT	village health team

OVERVIEW

Background

National programs and donor-funded projects increasingly rely on decentralized models of care to expand coverage of health services, ensure linkages to health facilities, and reach the most vulnerable populations. New emphasis has been placed on community-based models in which frontline health workers are expected to provide services and collect and report data. For example, the global "90-90-90"[1] targets recognize that achieving equity in HIV prevention and care will require an emphasis on community-based approaches and systems (Joint United Nations Programme on HIV/AIDS, 2014).

Community-level data can be helpful to health officials as they target programs and make decisions about care and services at lower levels of the health system. The United States President's Emergency Plan for AIDS Relief (PEPFAR) has embarked on a strategy to deliver the right types of interventions, in the right places, at the right time. This will require accurate, reliable, and timely data at district and subdistrict levels to provide an in-depth picture of community health so that programs can focus on populations most at need (PEPFAR, 2014).

Increasingly, community-based health programs collect data that flow into donor programs and national health information system(s) (HIS). Programs are turning to mobile health (mHealth) technology to address a variety of challenges. The mHealth tools provide solutions to challenges associated with paper-based reporting systems, such as inefficient filing systems and operational challenges including storage space associated with transporting paper forms and receiving data in a timely manner. Mobile technologies can help programs improve the completeness and accuracy of data, tap the potential for real-time reporting, and strengthen communication and supervisory feedback practices.

Community-based health information system(s) (CBHIS) should enable the sharing and use of information among community-based services and between community-based services and higher-level health facilities. Some of the data elements collected by a CBHIS should feed into the national health management information system (HMIS). mHealth can facilitate this process, because mobile devices can collect community data directly in an mHealth platform linked to the national HMIS. The use of mHealth varies depending on project scope, funding (government or donor), the state of HMIS expansion, and links with HIS. Below are examples of how mobile CBHIS are currently feeding into a national HMIS in three countries:

- **Zambia:** The National Malaria Control Program and PATH have a system where community health workers (CHWs) use low-cost mobile phones to collect information on malaria incidence (Roland, Sanner, Nilsen, & Braa, n.d.). A trained CHW then aggregates the information for a team of CHWs, and reports the data monthly into DHIS 2 using a Java-enabled mobile phone. A community mobile DHIS 2 platform is currently being tested.
- **Uganda:** The mTrac system is a mobile phone-based HIS-strengthening tool that uses text messages (SMS) to send data to the HMIS (mTrac, n.d.). The mTrac system is used by village health teams (VHTs) to submit weekly aggregate data (such as identified malaria and malnutrition cases, referrals, and drug stocks) through SMS based on existing integrated community case management (iCCM) registry and HMIS forms. Through the Saving Mothers Giving Lives pilot project, VHTs in four districts collect maternal health indicators on mobile phones and submit the information directly via SMS to the online national DHIS 2 system (DHIS 2, n.d.).

[1] By 2020, 90 percent of all people living with HIV will know their HIV status; 90 percent of all people with diagnosed HIV infection will receive sustained antiretroviral therapy; 90 percent of all people receiving antiretroviral therapy will have viral suppression.

- **Bangladesh:** The government provides laptops, tablets, and Internet access to community health clinics nationwide to support data collection and information sharing across the health system (Ahmed, 2012). Data entered in a laptop or tablet are stored in the central data bank, which is accessible to anyone with authorized login access to the health network. Those who provide health services will be able to see their clients' health profiles.

With the growing reliance on community-based health services to meet country and international health goals, the data from these services are being increasingly used by programs and governments at subnational levels to design interventions. Additionally, some governments and programs are decentralizing decision making to district and community levels. Information generated by community health services is key to informing decision making at all levels. In addition, CHWs use information on mobile devices to plan their home visits, provide health information to clients, schedule clients for follow-up, and track those who miss appointments. Given the power of community-based data to inform decision making, it is essential that these data are of high quality.

Conceptual Framework

The quality of data reported in a mobile data collection system depends on many factors. Two critical ones are highlighted in these guidelines: (1) design of the mobile data collection system and (2) implementation of data accountability protocols.

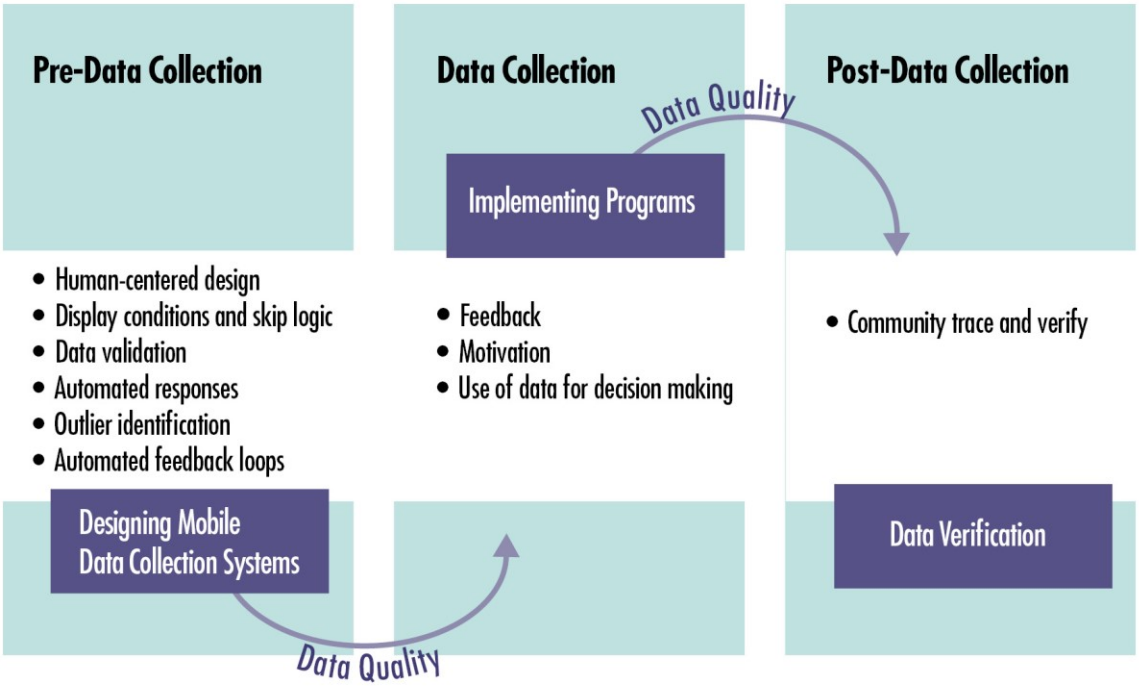
First, data quality is affected by the design of mobile data collection applications and platforms. Built-in components in the data collection workflow (such as skip logic, automated calculations, data validations, and instructional prompts) are used to increase the accuracy and completeness in data collection.

Second, data quality is affected by the feedback and supervisory structures in place for encouraging data collectors and their supervisors to collect and report high quality data. Data quality is affected by behavioral factors of the people collecting data, such as their attitudes, values, and motivation. This is increasingly important with a CBHIS, as data collectors generally are less educated and skilled and are dispersed across wider geographical areas with limited oversight.

In addition to well-designed data collection tools and formats, mHealth provides an opportunity to streamline work processes for data collectors and facilitate more real-time communication and feedback between data collectors and their supervisors. Greater communication and motivation among CHWs can lead to greater ownership of data and commitment to high-quality data.

Figure 1 presents the conceptual framework for these guidelines.

Figure 1. Conceptual framework for improving data quality in mobile community-based HIS



Objectives

The objectives of the guidelines for strengthening data quality in mobile community-based HIS are to help users in the following ways:

- Inform the design and implementation of a mobile CBHIS to collect, manage, and report quality data
- Assess the system's strengths and weaknesses
- Implement corrective measures with action plans to strengthen the system and improve data quality
- Monitor and track the system's capacity improvements and performance

Components of These Guidelines

These guidelines have four parts:

Part 1. Designing Mobile Data Collection Systems for Improved Data Quality. This component explains how to design a mobile data collection system and offers a checklist for assessing mobile data collection forms and systems.

Part 2. Implementing Programs to Increase Ownership of Data and Commitment to Data Quality and Data Use. This section explains how to engender accountability and ownership for data quality and offers a checklist for assessing feedback loops, supervisory structures for motivating CHWs on data quality issues, and systems and processes to promote the use of data for decision making.

Part 3. **Verifying Field-Level CBHIS Data.** This component of the guidelines explains how to adapt a community trace and verify (CTV) tool to verify whether people reported to have received services did so.

Part 4. Conducting a Mobile CBHIS Data Quality Assessment. This component explains how to determine the assessment's purpose, conduct community visits, and develop a system strengthening plan with follow-up actions.

Uses of These Guidelines

These guidelines are designed to be flexible. Countries and programs are encouraged to apply and adapt the guidelines according to their needs and contexts. The guidelines can be applied or components can be implemented separately according to programmatic requirements. The guidelines are useful for designing a mobile application and planning an mHealth program. The assessment can be applied routinely during program implementation.

Here are examples of the potential uses of these guidelines:

- **Initial and follow-up assessments of data management and reporting systems**. The assessment can be applied during the system design and planning phase and implemented routinely to identify gaps and monitor necessary improvements.
- **Routine data quality checks as part of ongoing supervision of frontline health care workers.** The assessment can be used to strengthen supervision mechanisms by assessing feedback loops and motivation for data quality.
- **Strengthening program staff capacity in data management and reporting**. Frontline health workers can be trained and sensitized to the domains linked to the collection and reporting of high-quality data.
- **Preparation for a formal data quality audit**. The guidelines can help identify data quality issues and weaknesses in the system that should be addressed before a formal data quality audit.

PART 1. DESIGNING MOBILE DATA COLLECTION SYSTEMS FOR IMPROVED DATA QUALITY

Introduction

The true power of community-based health data can be realized when stakeholders across all levels of the system use this information to make decisions that can improve health services and outcomes. Linking community data with the national HMIS can provide a comprehensive view of health progress and needs throughout the system, provided that the data are high quality. As community information is aggregated at higher levels, systems like the DHIS 2 have several data quality checks in place. However, as community information transitions from paper to mobile collection systems, best practices for data quality should be incorporated into the design and implementation before data are aggregated.

In this section of the guidelines, we focus on how mobile data collection applications and platforms can be designed to improve data quality. We provide guidelines for designing and creating mobile data collection systems and for organizations to assess themselves when designing mobile data collection forms and systems in the field.

Guidelines for Strengthening Data Quality in Mobile Community-Based Health Information Systems

BEST PRACTICES

- **Human-Centered Design**
- **Display Conditions and Skip Logic**
- **Data Validation Conditions**
- **Automated Responses**
- **Outlier Identification-Post Data Entry**
- **Automated Feedback Loops**

As mHealth solutions become more commonplace, new applications and platforms have emerged to help programs design mobile data collection systems for community-based health programs. Many applications or platforms include the data quality assurance mechanisms that are listed here, however they must be incorporated into the workflow by those designing the application to fit the relevant context. Our guidelines for best practices are meant for a wide audience ranging from those who are developing their own applications from the ground up, to those who are creating data collection forms using existing platforms.

Human-Centered Design

Designing mobile applications for improved data quality starts before the technology is introduced into the program. Human-centered design is a multi-stage problem-solving process that requires designers to analyze and foresee how users are likely to use a product or service (Rouse, 2007). Designers consider the needs, wants, and limitations of end users of the mobile data collection system at each stage of the process. Additionally, human-centered design tests the validity of designer assumptions about user behavior in real-world tests with actual users.

Principles for Digital Development (DigitalPrinciples.org, n.d.) incorporates human-centered design in the first principle: Design with the user. This means that solutions should be developed within the appropriate context with all user groups included in the design process. This ensures that the education and skill level of the data collectors are considered. Another important factor is device selection, considering users and their operating

DESIGN WITH THE USER

- Develop context appropriate solutions informed by user needs.
- Include all user groups in planning, development, implementation and assessment.
- Develop projects in an incremental and iterative manner.
- Design solutions that learn from and enhance existing workflows and plan for organizational adaptation.
- Ensure solutions are sensitive to, and useful for, the most marginalized populations: women, children, those with disabilities, and those affected by conflict and disaster.

Source: *DigitalPrinciples.org*

environment. For example, in circumstances where data collectors wear protective gear, such as bulky gloves, the data collection process should incorporate the use of a tablet with large buttons.

Additionally, solutions should be sensitive and useful for beneficiaries, such as marginalized populations. Projects should be designed in an incremental and iterative manner to allow for changes to be incorporated gradually. Ultimately, solutions should be designed to build on existing work flows to facilitate organizational adaptation. Involving users in the design process can result in a product built to strengthen and support user interactions. In these ways, using a human-centered design approach can result in improved quality of the data from CBHIS.

Human-centered design principles can help system designers understand how work flows are carried out, compared to how they are depicted in program documentation. This may involve observing the workflow as it is occurring or simulating the data collection process to incorporate the user perspective. Using this design process can reveal implementation steps that have been misunderstood or missed by the program. Identifying these discrepancies can improve data quality by updating the work flows (and data collection forms) to incorporate these changes and remove redundancies. Medic Mobile, a nonprofit organization specializing in mHealth, uses an interactive card-sorting process to engage users in role plays to enhance their understanding of work flows and how solutions can be designed.

Additional resources on human-centered design are listed in the References section.

More Resources

There are several existing resources describing the human-centered design process and how to incorporate these approaches when designing programs:

The mHealth Planning Guide provides tools for concept development, solution design, and testing, which can improve the quality of data collection. For more details and tools for developing, designing, and testing mHealth solutions for improved data quality, visit: https://www.k4health.org/toolkits/mhealth-planning-guide.

Principles for Digital Development provides resources on designing with the user in mind. For these resources on designing with the user, visit: http://digitalprinciples.org/category/principles/design/principle1-toolkits.

The Field Guide to Human-Centered Design is step-by-step guide to begin solving problems like a designer. The guide can be found here: http://www.designkit.org/resources/.

Display Conditions and Skip Logic

Mobile data collection systems are based on forms designed to guide the collection of the most relevant information. This is generally associated with a specific work flow. The forms display conditions and "skip logic" to determine when a given question should be used. By limiting the questions in the form solely to those users who will generate required information, the mobile data collector can narrow the focus and save time, compared

to paper-based data collection. Electronic forms with built in logic mandate that the correct data collection patterns will always be followed. Hence, reliance on the subjective logic of a data collector is eliminated.

In the context of CBHIS, the display conditions and skip logic used in mobile forms can also be configured to serve as a job aid guiding the data collector through a work flow. Under such a scenario, not only can data quality be improved, but the mobile tool can lead to improved quality of care providing decision support to the user, while also providing higher-quality data used for programmatic decision making.

Guidelines for Display Conditions and Skip Logic

Switching from paper-based forms to mobile-based forms does not mean simply creating a digitized copy of the paper forms. Paper and digital forms have different features and what works well on paper may not work well on a mobile platform. It also does not mean that digital forms will obviate the need for paper forms. The entire system may not go digital, and may still require paper forms in some cases. The need for a hybrid system using both paper and digital forms should be assessed, and tools (paper and digital) should be created accordingly.

Digital forms allow for design features that can facilitate data collection, but they require different thinking than paper forms. For example, when one enters an address on paper, it is customary to start with the detailed data (such as the street address) and continue to higher-level information (such as the province/state and country). However, when using a mobile platform, one starts from high-level data (country, province/state, and city) and proceeds to the more granular-level (street address), thus allowing the use of display conditions with predetermined options that can facilitate data entry. Table 1 provides guidelines for display conditions and skip logic.

Table 1. Guidelines for display conditions and skip logic

Display conditions and skip logic can be based on	Description	Example(s)
Characteristics of the user	Specific forms or information may only be relevant based on demographic information that has been previously entered. In this case, forms can be segmented based on the age, gender, or symptoms of a patient.	If the user indicates gender, there may be a different work flow for the patient. For example, if the patient is male, the pregnancy module can be skipped.
The time of year or date	Information may only be required or relevant during specific times of the year. Display conditions can be programmed to show only those fields required during specific time periods.	If a program is using CHWs to follow up on low coverage areas during vaccination or mass drug administration campaigns, workers may only present those screening questions during the period directly following those immunization campaigns.
The action required	When information previously entered requires an action, forms can be programmed to automatically appear and prompt the action.	If the user enters that a child has a fever over 104 degrees (F) the CHW is automatically notified to seek emergency care and the data collection will end.

Data Validation Conditions

Data validation conditions facilitate more accurate responses from users by requiring that inputs submitted meet criteria identified in the validation rule. Validation rules can be programmed in a variety of ways, from simply requiring a field to be entered before the form can be saved to requiring a specific type of response in a particular field. See Table 2 for examples. Validations improve data quality by preventing users from entering data that aren't logical or within specified parameters.

Data validation conditions on mobile forms can facilitate accurate data entry by providing help text that guides the user to enter an acceptable response. Such notification should be as detailed and clear as possible. For example, if the field requires a patient identification number that should be seven digits long and only include numbers, the notification text should define the acceptable value as "seven digits using numbers only." This enhances usability and promotes the correct response.

Data validation should be designed carefully, however, as it may prevent users from completing an entire form when they do not have an answer to one of the required questions or when they want to provide an answer in a format different to the one allowed by the validation rule. Moreover, while error messages or cell-specific help text are ideally employed to guide users, not all platforms allow for the display of such specific help messages. (For example, a form builder may only allow for the display of a single error/help message.) Some platforms also limit the number of characters that can be displayed in help messages, which may inhibit the amount of guidance provided to users. Increasingly, mobile data collection systems allow for the audio feedback to guide the data collector through data entry. Field testing of forms is essential to ensure that data validation is clear, appropriate, and constructive, and it does not impede data collection.

Table 2. Guidelines for validation conditions

Validation conditions can be based on	Description	Example(s)
Requiring questions to be submitted	Requiring responses for specific fields prevents important information from being skipped.	Requiring a patient identification number, facility number, or date of birth can provide important details needed for longitudinal tracking.
Requiring specific characters	When responses require only specific characters (such as alpha or numeric), fields can be programmed to only accept those characters.	Patient identification numbers may only be comprised of numbers, therefore entering an alphabet value would return an error.
Limiting open or free text	When possible, open-ended or free text responses should be limited, and instead the data collector should be provided via populated responses, drop-down menus, or lists with checkboxes. If multiple options are possible, creating validation relationships between options so that only logical combinations are possible to select can limit contradictory, inaccurate information.	When providing information on the location, the district is first selected from a drop-down menu rather than allowing for free-text entry. After the district is selected, the villages that are located within that district appear in the next drop-down option.
Requiring alpha or numeric responses within specified ranges	Alpha or numeric characters in a specific range can be programmed as a field validation. Numeric response can also be denoted as requiring values greater than or less than specific numbers.	Body temperature fields can include a validation restricting input from 95°F to 105°F or 35°C to 41°C. Patient age will likely not be greater than 120 years, which can be set as an upper limit.
Restricting the length of a string of characters	Validations can be set to limit the entry so that acceptable responses are restricted to a specific number of characters.	Depending on the country, phone numbers must include a specific number of digits to be valid. Validations can be programmed based on the acceptable values in the given context. See below for additional best practices for phone numbers.
A number or word entered following a specific format	When data entry should follow a specific format, the validation can be programmed to only accept values meeting the pattern criteria.	Standard email address formats require an "@" symbol, which can be required to ensure the contact details are recorded correctly. An ID system might require that the first two characters are alpha and the following four are numeric.
Date ranges	Date fields may require dates submitted are restricted to the current date or past dates. Date ranges may also be checked in chronological reference to other dates previously entered.	The first day of the last menstrual period is always entered as either the current day or a date in the past. A date of birth is not in the future or after a date of death.
Validating phone numbers	Depending on the country and type of phone, phone numbers can	When collecting phone numbers from beneficiaries for follow up, at

	be checked to verify if they are real.	the source of data entry, you can confirm if the number exists and therefore detect data collector typos, fake, or invalid phone numbers.[2]
Requiring a photo	When applicable, the application can prompt the user to capture a photo to verify information recorded.	When completing a household visit, the CHW could be required to capture a photo of the home, verifying that the visit occurred.

Table 3. Date validations
Required dates are opportunities to standardize entry in order to improve data quality and consistency.

Entering Dates:	Description:
Date format	Date format should be clearly labeled and relevant to the context. For example, requiring mm/dd/yyyy or dd/mm/yyyy.
Type of calendar	When the Gregorian calendar is not used in a specific context, options can be programmed to use different calendars. This should be clearly labeled on the date field to avoid confusion.
Entering dates	Dates can be selected from a calendar drop down rather than manually entering the month, day and year. To prevent the entry of certain date ranges (such as future dates), unacceptable dates can be formatted so that they are not able to be selected by the user.

Validation conditions can facilitate accurate responses, but there may be instances when users confront a novel situation that the software may not be prepared to accept. Providing an "other" field is an option that allows users to respond when a pre-determined response is not available; this can avoid the frustrations that users may experience when they cannot faithfully give a normalized response, especially if the software refuses to save their responses.

Automated Responses

When appropriate, fields may be programmed to automatically provide the value either by running a calculation or by using metadata that are automatically collected. Providing the values automatically can improve data quality by avoiding human calculation errors and saving time. Depending on the context, automatic calculations or metadata collection may be displayed to the users or only calculated/collected and stored in the information system.

Calculations

Calculable cells within a form can be programmed to be auto-filled based on data entry into previous related cells. Automatic calculations can limit the need for manual calculations completed by data collectors or when specific values may be unknown to the patient from which data are being collected. For example, clients may not know their exact date of birth, but they may know their approximate age. In this instance, an approximate birthdate can be automatically calculated from the age provided by the client. Conversely, age can be calculated if the date of birth is known.

Calculations can also be helpful when the health worker is supplying information to the patient, or recording information regarding the provision of follow-up services. For example, automatic calculations can be used to

[2] Information on code parsing, formatting, and validating international phone numbers can be found online at this link: https://github.com/googlei18n/libphonenumber

estimate the delivery date of a pregnant woman based on the first day of her last menstrual period. Automatic calculations can also be used to schedule future immunizations or delivery of medication.

Metadata

Many mobile data collection systems can automatically collect information that can be used in forms and for data quality analysis. Metadata are data that are not requested on the forms but that can be automatically collected. Metadata can be used to verify data quality and automatically record information about the data being collected. Here are some examples of metadata that can be automatically collected in forms:

- **User name (name of the person completing the form)** can be linked to user accounts within the system. For example, users may be asked to sign into their account before starting data collection. Once signed in, all data collected are linked to the users.
- **Date and time** can be collected automatically when the form is opened and saved. This can be pulled from the device settings. (Devices settings must be set up correctly.)
- **Location** can be collected via global positioning satellite (GPS) coordinates within mobile forms. If areas have been previously mapped, GPS can be used to passively collect location data thus obviating the need for data collectors to manually enter this information. In the case of passively collecting location data, users are often shown the location detected by the GPS system and asked to confirm it.

  Audio capture can randomly be activated to record a short portion of the interaction to ensure the interaction has occurred. Prior consent should be obtained from mobile users and patients.

Additional data can be collected regarding users' interactions and activities within the application that can be useful in verifying data and identifying potential outliers.

Outlier Identification after Data Collection

Depending on the mobile data collection platform, outlier identification may be built into the process. In other systems, this analysis may be run offline once data are exported to a CSV or an Excel file. Outliers can be identified through a self-normalizing approach based on standard deviations of previously collected data. Challenges with limited bandwidth and connectivity, however, may make it difficult to update the data collection platform with each version of data collection. Connectivity should be considered, to determine whether outlier identification is appropriate to run at the point of entry or after data have been transmitted. The following guidelines will be useful for planning how to run outlier identification.

Duplicative and conflicting submissions. These submissions can be detected when they are entered; programmed rules can be applied to determine a course of action. Duplicative data may require manual confirmation of unique identification fields before they are confirmed as genuinely duplicative. Once duplication is confirmed, if there is any difference between the two submissions of data, programmed rules or manual validation may be required to denote which duplicative entry to discard. Conflicting submissions should be reviewed via a structured process to determine which data are overridden. This decision could be based on time stamps, completeness, or manual validation procedures.

Completeness. This important component of data quality can be easy to assess and identify. Single data fields or entire forms can be missing.

> **Missing fields.** When fields are missing within a form or module they should be identified and analyzed further to understand the underlying cause. Missing fields could be due to omission from the user, or they could indicate an issue in the work flow of the application.

Missing forms. When entire forms are missing from specific modules or surveys it can be an indication of a user error or an error in the logic flow of the form. When missing forms are identified, feedback loops should be in place to understand the underlying cause, and the data collector should be notified.

Individual trends. When one is tracking longitudinal patient data, previous data associated with the patient can be used to identify trends over time. These trends can identify data that are not consistent with identified personal trends. For example, when one is tracking a child's body mass index (BMI) if a value were to significantly increase and deem the patient morbidly obese, the data can be flagged at the point of entry or after the information has entered the database.

Trends over time. These can be tracked to see if submitted data are inconsistent with identified historical trends. To establish yearly or seasonal trends, it may be necessary to analyze historical data from a non-programmatic source or to collect enough programmatic data to see if these trends can be detected.

Standard deviation outlier analysis. Analysis based on standard deviation provides a mechanism for revealing values that statistically diverge from previously collected data and thus may not be accurate.

Metadata. Data automatically collected from within the data collection system can be analyzed to determine outliers in data sets.

Time stamps. Tracking the start and end time associated with filling an electronic form allows for analysis of the total time spent completing the form. Completion time can be used as a predictor, or it can confirm if a form was falsified. Forms that are filled very quickly may indicate the data collector is not following the prescribed work flow or completing the required patient-health worker interaction (Birnbaum, 2012). Time stamps can also detect when a user is not conducting data collection as recommended or required by the protocol. For example, if a user is entering data at night, especially for several records, this could indicate that the user is not entering data at the point of encounter with the community member or patient as outlined in the protocol.

GPS coordinates. The location where information was collected can be recorded and analyzed using GPS map coordinates. This can help supervisors detect outliers, for example, when data collection is recorded outside of a health worker's catchment area or when multiple encounters are recorded at the same location. GPS coordinates are also an important data point that can be used in later data verification exercises. Supervisors or auditors can use the GPS location where the original data collection occurred to return to the location and verify data accuracy.

Another Resource

MEASURE Evaluation has developed many data quality resources, including guides and tools. For more information and links to these resources, please visit:
http://www.measureevaluation.orgrgcpc.unc.edu/measure/our-work/data-quality.

Automated Feedback Loops

Digital data collection systems provide an opportunity to introduce automated feedback loops that can provide actionable information to data collectors and supervisors to directly improve data quality or to improve the performance of the data collector. This can then improve data quality. Feedback loops provide insight into data being collected and allow for real-time detection of anomalous data or other data quality issues. Depending on connectivity, real-time data monitoring can allow supervisors to monitor data as they are collected. Monitoring data in real time allows supervisors to spot and remedy problematic data as issues occur. Here are a few examples of automated feedback loops:

SMS messages. Systems can be programmed to use short message service (SMS) messages that automatically notify the data collector or supervisor when data are flagged for quality concerns or data are missing. This allows issues to be identified and remedied as they arise.

Dashboards. Dashboards are data visualization tools that track key performance indicators, allowing for easy monitoring of data to glean insights for decision making. With data available in real time, dashboards can be configured for use by supervisors, data collectors, or other community-based data users. Dashboards can be designed to track performance and program indicators in addition to monitoring data quality. This helps data users understand data issues at a macro level so that appropriate responses can be designed. Color indicators (such as the red-yellow-green spotlight system) can be effective to provide intuitive interpretation of data.

Notifications to submit reports or forms. Automatic notifications can be set up so that data collectors are reminded to submit their data, thus saving supervisors' time. When reports are missing, notifications can automatically be sent to remind the data collector to submit the report. If data collectors fail to respond to the report reminder, it may then go automatically to the supervisor for follow-up action.

Feedback when form is submitted. After a data collector has submitted a form, a confirmation of its receipt and recording can be sent back automatically via SMS message or other automatic notification. When forms are saved on the device without submitting to the system, notifications can be sent automatically to prompt the user to submit the form when connectivity is available.

After receiving feedback about an outlier or quality issue, verify that appropriate steps were taken to correct it.

Table 4. Checklist for designing mobile data collection systems for improved data quality

Human-Centered Design[3]	
□	Solutions are context appropriate and informed by user needs.
□	All user groups are included in planning, development, implementation, and assessment.
□	Projects are designed in an incremental and iterative manner.
□	Solutions are designed to learn from and enhance existing work flows and plan for organizational adaptation.
□	Solutions are sensitive to, and useful for, the most marginalized populations: women, children, those with disabilities, and those affected by conflict and disaster.
□	Data capture tools are pilot-tested in the field with all user groups.
Display Conditions and Skip Logic	
□	When data collection forms are switched from paper to digital process flows are carefully considered and display conditions are built in to enhance existing workflows.
□	Forms and questions are configured to include display conditions and skip logic patterns.
□	Only those forms or questions relevant to the patient are displayed by configuring display conditions and skip logic.
Data Validation	
□	Program-critical fields (such as patient ID) are programmed to require a response.
□	When relevant, validation rules are configured so that data fields only accept alpha or numeric characters.
□	Drop-down menus or check boxes are provided to avoid the submission of open ended free text.
□	Drop-down menus and check boxes are designed to only display logical combinations based on the previously entered fields.
□	When relevant, validation rules are designed to require numeric entries within a pre-specified range.
□	When dates are entered, validation rules require formats to be standardized and entries are restricted to the relevant period.
□	When phone numbers are entered, the format of the phone number is standardized and validated.
□	Validation rules are designed to crosscheck data entered in previous fields to restrict contradictions (such as age and DOB).
□	Helpful text or guidance (for correction) is provided when data is entered that does not meet a validation rule.
Automated Responses	
□	When calculations are required, the system is configured to automatically calculate values based on previously entered data.
□	Metadata like timestamps and GPS coordinates are automatically captured from the device.
□	When possible, metadata automatically captured is used in place of requiring fields to be completed by the user.
Outlier Identification	
□	A mechanism is provided to identify outliers or errors after data has been entered and submitted.
□	Duplicate data submissions are detected, confirmed, and deleted based on time stamps, completeness, or manual validation procedures.
□	A mechanism is provided to identify missing forms and/or fields.

[3] The human-centered design checklist items are adapted from DigitalPrinciples.org (n.d.).

□	A mechanism is provided to track trends in data across time. This can be done at different units of analysis (such as for one person or one geographical catchment area).
□	A mechanism is provided to identify outliers based on standard deviation analyses.
□	Data errors and outliers are identified using metadata (such as time stamps or GPS coordinators).
Automated Feedback Loops	
□	End users are notified upon successful submission of data.
□	Messages are directly sent (through SMS or email) to the data collector or supervisor when a data quality issue is identified.
□	Reminder messages are directly sent (through SMS or text) to data collectors to submit their data and reports.
□	Supervisors can monitor data in real time through dashboards or other mechanisms so that data quality is monitored and performance is tracked.

PART 2. IMPLEMENTING PROGRAMS TO INCREASE MOTIVATION AND COMMITMENT TO DATA QUALITY

High data quality requires CHWs who are motivated, committed, and equipped to collect and report accurate data in a timely and effective manner. Mobile health tools can improve accountability for data quality by providing CHWs with helpful feedback as the data collection and reporting loop is completed and problems are identified and remedied effectively. These tools also can motivate CHWs to provide high-quality data for use by decision makers at all levels. The guidelines and checklist presented in this section are designed to increase the motivation and commitment of CHWs by equipping them with mechanisms to ensure data quality. Some strategies identified can be used on a variety of performance metrics, but for the purposes of these guidelines we focus on data quality.

Guidelines

Feedback

Mobile technology tools can be used to provide data collectors and their supervisors with immediate, direct feedback on data quality. Identifying and responding to data quality issues up front—at the point of data collection—significantly decrease the likelihood of data entry errors, and save the time and money that would be associated with revisiting and cleaning data errors later. However, mechanisms are needed to respond to feedback messages on data quality quickly to close feedback loops effectively. mHealth tools can assist supervisors and CHWs by tracking data quality performance, communicating data quality issues, and linking to action steps that must be undertaken to improve or rectify identified data quality problems. The following guidelines are designed to facilitate effective feedback mechanisms to enhance data quality.

- **Provide supervisors with regular reports tracking the data quality performance of their teams.** Regular targeted supervisory reports (such as tables and charts on key data quality performance indicators in paper reports or mobile dashboards) should be made available to supervisors. For example, D-tree International uses customizable supervisory dashboards that allow for real-time tracking of key performance and data quality indicators on phones, tablets, and computers (D-tree International, n.d.). Supervisory reports can be used to monitor the data quality performance of teams and alert supervisors to suspicious data that should be followed up directly with health workers and their beneficiaries.
- **Tie data quality alerts/flags to actionable follow-up items that can be tracked.** Alerts about suspicious or inconsistent data should always be accompanied by actionable tasks to improve data quality that must be followed up and tracked. Follow-up tasks to improve data quality can take place during regular supportive supervision visits, data quality review meetings, or administrative visits. Tracking follow-up tasks can occur in many ways, such as logging CHW attendance at data quality review meetings, documenting spot checks with program beneficiaries, and logging time of meetings with health workers. This can build accountability and ensure that supervisors and health workers are following through on their responsibilities.

  For example, Malaria Consortium's inSCALE project used a mobile supervisory CommCare application to track and address CHW performance issues. This application tracked health workers' data submissions and benchmarked indicators, reporting rates, and timeliness of reports across members of their teams. The application also incorporated active data management capabilities that provided supervisors with automated timely reports and targeted follow-up action steps when issues were identified among their teams (Dimagi, 2014). Automated messages flagged problems with submitted

data (such as timeliness and completeness), and alerted supervisors about which CHWs they should contact to provide more targeted supervision (Malaria Consortium, 2014).

- **Provide CHWs with timely, useful, and supportive feedback on the quality of their submitted data.**
  Feedback on data quality should include appropriate information about the accuracy, completeness, and timeliness of their reports. Messages that acknowledge the achievements of CHWs can be motivating, making them feel supported and valued.

  Traditionally, there was only a one-way flow of data collected by CHWs and reported up to facility, district, regional, and higher levels of the health system. Community health workers often receive little to no feedback on their performance, which can be demotivating and enhance feelings of isolation among workers (Thondoo, et al., 2015). Mobile tools can be used to provide corrective and supportive feedback to encourage high CHW performance. Corrective feedback that indicates areas needing improvement or corrective action can be prompted in a mobile system through automated fault messages explaining the data quality error, along with a request to resubmit. Supportive feedback can reinforce current behaviors and skills by identifying what is being done well. Such positive reinforcement is especially important for CHWs, who as low-paid staff or volunteers may need validation and encouragement to feel supported and valued. To encourage good practices, motivating messages can be automated and sent to CHWs who send complete, high-quality data on time.

  Mobile (mHealth) tools can facilitate more direct and regular communication with community data collectors on the quality of their data. Mobile tools can be used to provide real-time feedback to health workers via web-based tools, dashboards, messages, or voice-based tools. For example, the ASHA Self-Tracking Application (ASTA) was a web-based and phone-based system developed for accredited social health activists (ASHAs) in India (DeRenzi, et al., 2016). This tool displayed simple visualizations of ASHA performance compared to a subset of their peers, progress against personal targets, and a comparison of current performance to past performance. Similar tools can be used for data quality performance. Personalized tools such as this provide a mechanism for CHWs to see the immediate use of the data they are collecting and allows people to track their own performance.

- **Institute regular (monthly or quarterly) meetings where CHWs discuss data quality and are engaged in developing appropriate solutions to rectify identified issues**.
  Feedback on data quality should also be included in routine supportive supervision visits. Regardless of the mechanism, feedback systems should be developed in a participatory design process to ensure that the format and content of the feedback is relevant to CHWs' regular work responsibilities, and appropriate for their capacity and literacy levels.

- **Develop a documented protocol for higher-level data quality checks**.
  This should involve regular mechanisms for advanced checks (such as data quality spot checks) to detect suspicious patterns, inconsistencies, or anomalies in data. Checks may be automated, but some may require manual review of the data. Manual checks should consider the context of the program; these may be undertaken at the managerial level and require additional training in data analysis skills.

Motivation

Users of mHealth data collection tools are often CHWs who also provide patient care and services. The additional responsibilities of primary data collection and reporting can be burdensome for workers who volunteer or receive low pay. Health workers who understand the importance of data are more motivated to ensure its quality. The following guidelines are designed to enhance CHW motivation.

- **Highlight the value of data quality at the start of recruitment and training processes for CHWs**. Emphasis should be placed on the value of data quality and its role in providing useful information to the community. During the recruitment and training processes, education campaigns can be used to emphasize the importance of data quality and engage CHWs in data quality processes, so that CHWs who are accountable for data collection understand that ensuring high data quality is part of their routine roles and responsibilities.

- **Make comparative data quality performance data available to health workers, supervisors, and teams to promote competition and motivate improved performance.**
  Rankings of health workers or teams can be provided to supervisors or CHWs on a regular basis via paper or mobile reports. Dashboards can also be created that track a CHW's data quality performance against programmatic data quality goals (such as 90 percent of reports submitted on time) or against the performance of other health workers. Depending on the needs of the program and what works best to drive performance, comparisons can be made against other CHWs in a team, other groups of CHWs, or the highest performing CHW in their group. This data can be aggregated at a cadre level to facilitate comparisons on data quality against other teams, catchment levels, or districts. For example, Catholic Relief Services (CRS), Dimagi Inc., and Vatsalya implemented the Reducing Maternal and Newborn Deaths (ReMiND) program. Government CHWs used basic mobile phones operating Dimagi's open-source CommCare software to support client assessment, counselling, and the identification, treatment, and referral of pregnancy, postpartum, and newborn complications. As part of this platform, the ASHA Self-Tracking Application (ASTA) allowed frontline health workers to compare their performance and ranking against a subset of their peers; this was self-motivating and built confidence (DeRenzi, et al., 2016). Promoting competition against teams can also help foster accountability, connectedness, and cohesiveness among group members as they encourage each other to improve data quality and increase team performance.

- **Facilitate communication between team members and supervisors.**
  This can improve CHWs' motivation to their work by reducing their feelings of isolation and frustration due to remote placements. Closed user-group networks among team members and their supervisors can be used to share helpful advice and lessons learned about data quality or to send reminders to submit complete reports on time. For example, the Millennium Villages Project in Malawi uses WhatsApp, an instant messaging application for smartphones, to connect CHWs and their supervisors, enabling them to disseminate information on performance, seek advice from others, solve problems, acknowledge performance, motivate each other, and share empathy. Preliminary assessment findings show that CHWs greatly appreciate the increased speed and quality of communication within their teams. The groups have served as promising mechanisms for enhancing professional connectedness (Pimmer, 2015). Creating such supportive environments can help health workers learn from each other as they troubleshoot data quality issues.

  Sending unidirectional regular performance-related messages to CHWs can be a way to maintain motivation and encourage high data quality. For example, a study examining the use of supportive SMS was shown to increase the documentation and reporting of pregnancies by health surveillance assistants in Malawi (Joos, et al., 2016). These types of messages can be designed to appreciate CHWs ("Thank you for your contribution"), remind them about deadlines and reporting ("Remember to complete your quarterly report"), or tell them to focus on targeted data goals ("Remember to document stillbirths in your pregnancy report").

- **Implement a system of incentives to motivate high data quality performance**.
  Incentives can be purely financial (pay for performance), in kind (e.g., top up phone airtime), or recognition (e.g., certificate of achievement). Both financial and non-financial rewards can be a powerful motivator for increasing data quality performance.

Performance-based incentive schemes may be set up for health workers and their supervisors based on data quality performance, such as for reports submitted on time. For example, a malaria program implemented by PATH MACEPA, the Zambia Ministry of Health, and Akros Inc., rewards CHWs for timely data reporting by giving them additional mobile phone airtime or small financial incentives (Mendoza, Okoko, Konopka, & Jones, 2013). The workers receive the full amount of the incentive payment if they submit their reports on time, and only a fraction of the amount if they report late. The consistency of financial incentives is important for sustaining motivation; once in place, changing or removing incentives can demoralize workers.

Team-based financial incentives can be used to leverage peer support and friendly competition between teams. For example, small nominal financial incentives can be given to each member of the CHW team who ranks highest on a data quality performance indicator. Using a collective group model of rewards helps to build accountability across teams, because people may perform better to avoid letting down teammates. This model also could motivate people who can learn best practices from high performers on their team. Team incentives can build connectedness among team members as they work toward a data quality goal together.

- **Recognize CHWs within the communities where they live and work**.
  Receiving praise, respect, and validation from the community can be motivating. This can occur at public events where certificates, awards, or other forms of appreciation are given to CHWs who meet their data quality targets or have helped others solve data quality issues. Public recognition may be especially validating when it comes from respected community leaders or decision-making groups, providing a mark of prestige and status to the community health worker (African Strategies for Health, 2013).

Use of Data for Decision Making

Data quality and data use are intrinsically linked. Data quality naturally improves as people and organizations understand and value how useful data can be for program planning and decision making. The regular demand for data also depends on the availability of accurate, timely, and complete data that allow users to have confidence in the quality of information on which decisions are based. For CHW programs, strengthening service delivery and increasing access to care depend on the use of quality data to plan, monitor, and improve program performance. Here are some guidelines to facilitate the use of data for decision making:

- **Data from mobile CBHIS should be readily available in accessible and interpretable formats** to support the use of information in decision making, either through the mobile system or through other types of information products (such as graphs, charts, maps, and bulletins).
- **Data should also be in accessible formats for different target groups.**
  These include both program managers and decision makers—who may use information for program planning, monitoring, and improvement or policy review and development—as well as CHWs who can use data at the community level for service delivery through micro planning, day-to-day management of tasks, and performance monitoring. The different information needs of various groups should be considered to ensure that the correct information is targeted for communication and the appropriate level of detail and complexity of information is presented.
- **Data should be shared with community members during forums in which information about community health status and needs and program performance is shared with beneficiaries.**
  This can promote dialogue, feedback, and accountability for improved service delivery and program

quality. The format, complexity, and means of data communication and visualization should consider the role of the user, their education level, and their information needs. At the same time, capacity building efforts should focus on enhancing the data literacy and data use skills of CHWs.

Motivating high data quality performance from CHWs is intrinsically linked with establishing a culture of data use and ensuring that all health workers recognize the value of data. Understanding that others have a stake in the data collected and that the use of data can lead to tangible improvements in their work or in the health of their communities, can engender commitment to high data quality.

Communicating how data have been used at all levels to improve programs and health outcomes, after it has flowed through different levels of the health system, can motivate health workers to ensure high-quality data. Similarly, disseminating synthesized data or information back down the data chain to lower levels can help CHWs understand the status of their catchment area compared to other geographic locations. This helps them understand how their work contributes to program performance at an aggregated level.

Programs should place a high value on data through their structures and processes. Data-informed decision making must be clearly emphasized in policies, guidelines, and standard operating procedures. Linking data quality with data use for improvement during performance review meetings, regular work planning sessions, and ongoing supportive supervision visits can help emphasize the programmatic importance of both data quality and data use to track progress toward targets, identify issues of concern, and problem solve to improve performance.

MEASURE Evaluation has developed a suite of online data demand and use tools, with guidance documents, training, and assessments that can be adapted for use with mobile CBHIS (see MEASURE Evaluation resources in the References section).

Table 5 provides a checklist for programs to follow to increase CHW motivation and commitment to data quality

Table 5. Checklist for implementing programs to increase motivation and commitment to data quality

Feedback	
□	Information on the data quality performance of teams is made available to supervisors on a regular basis.
□	Data collectors receive regular feedback on the quality of their submitted data.
□	Data collectors are actively engaged in developing appropriate solutions to address data quality issues.
□	The format and content of data quality feedback has been designed in a participatory process, is relevant to CHW responsibilities, and is appropriate for the level of capacity and literacy of the user.
□	Feedback on data quality requires an actionable follow-up task that must be documented when complete.
Motivation	
□	The importance of data quality is widely communicated during recruitment and training of data collectors.
□	Comparative data quality performance is made available and accessible to supervisors and data collectors.
□	Mechanisms exist for data collectors, peers, and supervisors to communicate and support each other on data quality issues (such as peer support groups, closed network user groups, etc.).
□	Financial or non-financial incentives are provided to data collectors and/or their supervisors for high data quality performance.
Use of Data for Decision Making	
□	CHWs receive analyzed data (information products, charts, graphs, and maps) on a regular basis.
□	Analyzed data or data visualizations are tailored to the capacity level of CHWs.
□	CHWs have received appropriately targeted training on data interpretation and data use.
□	Guidelines exist to support the analysis, presentation, and use of data at community levels (such as interpretation guidelines for visualizations produced by mobile systems).
□	CHWs have access to guidance or technical assistance on data use (such as data review meetings or during supervisory visits).
□	Analyzed data/results are regularly presented/disseminated to a wide variety of stakeholders (other than data producers) in a timely manner so that information can be used to inform decisions.
□	Service delivery/programmatic decisions are made or corrective actions are taken by CHWs based on analyzed data.
□	CHWs regularly participate in data review meetings that discuss data quality and key program indicators with program managers and other decision makers.
□	Data quality is regularly discussed during routine supportive supervision visits.

PART 3. VERIFYING FIELD-LEVEL CBHIS DATA

Introduction

Data quality reviews and assessments for paper-based data collection systems are well-covered topics. As programs transition to data collection methods using mHealth tools, there is no paper trail to facilitate a quantitative comparison of recounted registers/reporting forms to reported data.

With the rise of mobile data collection at the community level, it is necessary to examine how data quality assessment (DQA) guidelines and tools can be adapted for this scenario. We recommend adapting the Community Trace and Verify Tool as a method for verifying field-level data. This involves tracking primary data that has been collected at the individual or household level.

Community Trace and Verify Tool

The CTV Tool was developed to provide programs with a way to verify whether orphans and vulnerable children (OVC) reportedly being provided with services by community-based organizations did so. The CTV questionnaire was not designed to be an inventory of all services received by OVC, but rather of selected services that can provide an indication of whether a program is reaching its target beneficiaries with services. The questionnaire has been kept short and its administration is not burdensome. The tool consists of three parts: questionnaire, tabulation plan, and implementation protocol.

The premise of the CTV tool is to visit beneficiaries to verify reported data for any program delivering services at the community level. Household visits or phone calls are made to administer a short 10- to 20-minute survey to verify whether beneficiaries received the reported services.

The CTV uses lot quality assurance sampling (LQAS)—a rapid survey sampling method that allows one to determine whether performance at a cluster level (CHW catchment area or supervision area) is acceptable (pass) or not acceptable (fail). In LQAS, a relatively small sample of households is surveyed in each cluster. A sample size of 19 households per cluster is considered acceptable (Davis, Luna, Rodriguez, & Sarriot, 2009).

Due to difficulties finding locations of households and household members to interview, we recommend sampling 40 households to reach the target of 19. The LQAS method is not meant to assess people's performance; rather it can be used to draw conclusions about performance at the level of a supervisory area or a larger program area. For the purposes of the mobile CBHIS DQA, the CTV tool uses LQAS to assess the accuracy of data collected (rather than the quality of service delivery performance).

This method of data verification differs from the usual paper-based recounting and comparison of reported data to raw data on registers or paper forms. Instead, the CTV assesses the accuracy of data recorded in the mobile health system, comparing "raw" data entered in a mobile system to reported data gathered from re-interviewing program beneficiaries. Care must be taken to minimize recall bias associated with this approach. Here are some strategies to minimize recall bias:

- Minimize timeframe for follow-up to allow for a short recall time (monthly or quarterly, as opposed to six months after a community health worker visit).
- Ensure interviewers are well trained and skilled to frame questions aid accurate recall.
- Select indicators that represent the delivery of a service or distribution of a commodity.
- Triangulate with other sources of data at the household, if they exist (such as household health cards).

Guidance on the purpose and implementation for the CTV tool (developing a protocol for field work; creating a template for data collection and analysis; and considerations for training, data collection, and feedback) is

available on the MEASURE Evaluation website (see References section). In this section, we describe steps to consider in adapting and implementing the CTV tool for mobile CBHIS data quality assessments.

Selecting Indicators

The selection of indicators for the CTV tool will depend on the type of community-based program being assessed. If possible, the indicators selected should represent the delivery of a service or distribution of a commodity, as these events are less susceptible to recall bias. For example, an iCCM program using mobile devices for data collection might use the following indicators for a CTV tool:

- Received a household visit from a CHW
- Received a mid-upper arm circumference (MUAC) tape test to screen for acute malnutrition
- Received a rapid diagnostic test (RDT) to diagnose malaria
- Received artemisinin-based combination therapy (ACT) for the treatment of uncomplicated malaria
- Received anthelmintic drugs to treat intestinal parasites
- Received oral rehydration salts (ORS) and zinc to treat diarrhea
- Received antibiotics (e.g., amoxicillin) to treat nonsevere pneumonia
- Received a referral to a facility

Depending on the type of CHW program, each household may receive different types of services. For each indicator chosen, 40 households that are recorded as receiving that service will need to be sampled from the mobile data collection system. Depending on the number and type of indicators selected for the CTV, there may be several sampling lists each with a specific questionnaire associated with the service.

Choosing a Pass/Fail Threshold

For each indicator, or for a set of indicators, one must set a pass/fail threshold in advance. This threshold will determine whether a cluster of households passes or fails. The threshold one sets depends on the level of data quality a program is attempting to achieve at a given point in time. For example, one might choose a lower threshold at the start of implementation of a CBHIS or with an indicator that might be subject to recall bias. Thresholds for indicators can change throughout the course of program implementation or scale up as data quality improves. Generally, for data quality purposes, we recommend a pass/fail threshold of 80 percent or higher. If a program wants a coverage target of 80 percent to say that they have strong data quality, then 13 of 19 households in each supervision area need to respond "yes" to the survey questions.

Selecting Number of Coverage Areas

The number of coverage areas selected in a CTV for a mobile CBHIS data quality assessment depends on program context, structure, and available resources. The LQAS typically treats a "supervision area" as a coverage area; the resulting data from an LQAS can be used to guide supervisory efforts in each individual supervision area and can be pooled in an estimate of performance across an entire program area.

Supervisory structures vary across community-based programs. For example, a CHW program in Malawi was designed so that one supervisor oversees a cell which is comprised of seven villages. Each village has 2–4 CHWs who are responsible for 40–60 households. In this case, one supervisory area could include more than 1,000 households. Sampling 40 households out of this cell would be labor-intensive and could lead to imprecise

estimates of data quality performance. Methods have been developed to adapt the LQAS sampling method across many strata (Hedt-Gauthier, Mitsunaga, Hund, Olives, & Pagano, 2013).

The selection of coverage areas for a DQA may be random or purposive. Random selection might be used for periodic (baseline and endline) assessments of data quality to obtain an overall impression of data quality performance. Purposive selection might be used to focus attention on specific coverage areas identified as exhibiting data quality issues through other aspects of the data quality assessment (such as those with a high level of flagged validation or outlier identification issues or those with poor program performance). These areas can then be targeted for intense supervisory practices and quality improvement activities.

Identifying Households to Interview

A full list of program beneficiaries should be generated from the mobile platform for each coverage area to be assessed. From this list, 40 households should be selected randomly for in-person visits or phone calls to verify information that has been collected through the mobile platform. Households should be oversampled to account for the mobility of populations; people may have moved or may not be home at the time of the interview. The LQAS method calls for responses from a maximum of 19 sampled households per coverage area. However, sampling can stop once the pre-selected pass/fail threshold has been met. Once data have been collected from 19 households or a pass/fail threshold has been met, fieldwork for the indicator in the coverage area is complete.

PART 4. CONDUCTING A MOBILE CBHIS DATA QUALITY ASSESSMENT

Determine the Assessment's Purpose

Designing systems and programs to collect high-quality data from the start can save time and resources in data quality audits or other verification processes after data collection. The first step is to determine the purpose of the assessment.

The mobile CBHIS data quality assessment can be used for the following purposes:

- Initial assessment of newly established mobile CBHIS. This assessment can be followed up routinely to monitor improvement plans.
- Routine ongoing supervision of data management and reporting systems.
- Periodic mobile CBHIS assessment by donors.
- Preparation for a formal data quality audit.
- External assessment of the data quality assurance mechanisms employed by a mobile CBHIS.

Conduct Community Visits

Once the purpose has been determined, the next step is to plan and conduct community visits to implement the assessment. During visits with mobile CBHIS program implementers, both document reviews and interviews with program staff will be carried out to complete the relevant sections of the assessment in the Excel file (Part 1, Part 2, and Part 3). The assessment team may consist of two or more assessors, depending on the size and purpose of the assessment.

The size of the program will determine the number of community visits one should conduct. Random sampling techniques can be used to select a representative group of sites in which data quality is indicative of data quality for the entire program. One might also want to compare assessment findings across various factors (such as geographical regions or implementing partners) and sample appropriately for this.

During community visits, the assessment team should observe the implementation and use of the mobile CBHIS and talk to CHWs to observe them interfacing with the mobile device to assess them against the criteria in the guidelines. Additionally, key informants should be interviewed (such as mobile application developers, field engineers, service designers, project deployment managers, mHealth specialists, CHW managers, and monitoring and evaluation officers and staff).

Staff should be notified prior to the visit for the DQA to ensure the appropriate personnel are available to answer the questions in the assessment checklist.

Develop a System-Strengthening Plan with Follow-Up Actions

After the system and the program are assessed, an action plan should be created based on the findings of each checklist. This should be done in collaboration with the team members who are involved in the data management and reporting process to create ownership of the plan and get direct insights on the feasibility and applicability of suggested interventions. Ideally, the team should review the checklist findings and lead the process of developing the action plans.

REFERENCES

African Strategies for Health. (2015). *Community health worker incentives in Malawi: Lessons learned.* Arlington, VA: Management Sciences for Health. Retrieved from https://www.msh.org/resources/community-health-worker-incentives-in-malawi-%E2%80%8Blessons-learned

Ahmed, S. (2012). Community clinic health revolution in Bangladesh. Dhaka, Bangladesh: World Health Organization/Bangladesh.

Davis, R., Luna, J., Rodriguez, A., Sarriot, E. 2009. *Rapid health surveys: Principles and sampling design handbook.* Calverton, MD: ICF Macro. Retrieved from http://robbresearch.pbworks.com/f/Rapid+Surveys+Handbook_Dec2009_final.pdf

Birnbaum, B.E., DeRenzi B., Flaxman A.D., & Lesh N. Automated quality control for mobile data collection. ACM Symposium on Computing for Development, 1-1, doi: 10.1145. 2012. Retrieved from http://dl.acm.org/citation.cfm?id=2160603

DeRenzi, B., Wacksman, J., Dell, N., Lee, S., Lesh, N., Borriello, G., & Ellner, A. (2016.) *Closing the feedback loop: A 12-month evaluation of ASTA, a self-tracking application for ASHAs*. Paper presented at International Conference on Communication and Technologies and Development (ICTD) 2016, Ann Arbor, MI, June 3–6, 2016. Retrieved from http://dl.acm.org/citation.cfm?id=2909652

DHIS 2. (n.d.). DHIS 2 in action. Uganda. Retrieved from https://www.dhis2.org/inaction

DigitalPrinciples.org. (n.d.). Principles for digital development. Digital Development Principles Working Group. Retrieved from http://digitalprinciples.org/wp-content/uploads/2015/05/Principles-Overview.pdf

Dimagi. (2014). *Malaria Consortium inSCALE: ICCM for improving child care*. Cambridge, MA: Dimagi, Inc. with London School of Hygiene and Tropical Medicine and University College of London. Retrieved from http://www.dimagi.com/malaria-consortium-inscale-iccm-for-improving-child-care/

Hedt-Gauthier, B.L., Mitsunaga, T., Hund, L., Olives, C., & Pagano, M. (2013). The effect of clustering on lot quality assurance sampling: A probabilistic model to calculate sample sizes for quality assessments. *Emerging Themes in Epidemiology,* 10 (1). Retrieved from https://link.springer.com/article/10.1186/1742-7622-10-11

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2014). 90-90-90: An ambitious treatment target to help end the AIDS epidemic. Geneva, Switzerland: UNAIDS. Retrieved from http://www.unaids.org/en/resources/documents/2017/90-90-90

Joos, O., Silva, R., Amouzou, A., Moulton, L.H., Perin, J., Bryce, J., & Mullany, L.C. (2016). Evaluation of a mHealth data quality intervention to improve documentation of pregnancy outcomes by health surveillance assistants in Malawi: A cluster randomized trial. *PLoS ONE*, 11(1), e0145238. Retrieved from https://jhu.pure.elsevier.com/en/publications/evaluation-of-a-mhealth-data-quality-intervention-to-improve-docu-4

Malaria Consortium. (2014). *Mobile health (mHealth) approaches to improve motivation and performance of CHWs in Mozambique.* London, UK: Malaria Consortium. Retrieved from http://www.malariaconsortium.org/resources/publications/448/mobile-health-mhealth-approaches-to-improve-motivation-and-performance-of-chws-in-mozambique

Mendoza, G., Okoko, L., Konopka, S., & Jones, E. (2013). *mHealth compendium*, Volume Three. Arlington, VA: African Strategies for Health project, Management Sciences for Health.

mTrac. (n.d.). How mTrac works. Kampala, Uganda: mTrac. Retrieved from www.mtrac.ug/content/how-mtrac-works

Pimmer, C. (2015). Using WhatsApp groups to support community health workers. Preliminary insights from a pilot study in rural Malawi (Interim Report). Retrieved from http://www.christoph.pimmer.info/wp-content/uploads/2015/02/WhatsApp-Groups-to-support-Community-Heatlh-Workers-interim-report1.pdf.

Roland, L.K., Sanner, T., Nilsen, P., and Braa, K. (n.d.). Mobile community reporting. Using DHIS 2 Java clients for Malaria community reporting in Zambia. Oslo, Norway: University of Oslo. Retrieved from https://www.dhis2.org/mobilereporting

Rouse, W.B. (2007). *People and organizations: Explorations of human-centered design*. Hoboken, NJ: John Wiley & Sons. Retrieved from http://onlinelibrary.wiley.com/book/10.1002/9780470169568.

Thondoo, M., Strachan, D.L., Nakirunda, M., Ndima, S., Muiambo, A., Kallander, K., Hill, Z., & the InSCALE Study Group. (2015). Potential roles of mHealth for community health workers: Formative research with end users in Uganda and Mozambique. *Journal of Medical Internet Research mHealth uHealth*, 3(3): e76. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26206419

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2014). PEPFAR 3.0. *Controlling the epidemic: Delivering on the promise of an AIDS-free generation*. Washington, DC: Office of the U.S. Global AIDS Coordinator. Retrieved from https://www.pepfar.gov/documents/organization/234744.pdf

MEASURE Evaluation Resources

Data demand and use tools. https://www.measureevaluation.org/resources/tools/data-demand-use/data-demand-and-use-strategies-and-tools.html

Data quality resources, guides, and tools. http://www.cpc.unc.edumeasureevaluation.org/measure/our-work/data-quality

The original tool, focused on orphans and vulnerable children: https://www.measureevaluation.org/resources/publications/ms-13-63/

Community trace and verify tool: https://www.measureevaluation.org/resources/webinars/community-trace-and-verify-in-tanzania

MEASURE Evaluation
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, NC 27517 USA
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) under terms of Cooperative Agreement AID-OAA-L-14-00004 and implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this presentation do not necessarily reflect the views of USAID or the United States government. TR-17-182

ISBN:978-1-9433-6459-6

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE
Evaluation