VillageReach is a global health innovator that develops, tests, implements and scales new solutions to critical health system challenges in low-resource environments, with an emphasis on strengthening the "last mile" of healthcare delivery.

www.villagereach.org // info@villagereach.org
# Table of Contents

Table of Contents ................................................................. ii

ODK Scan Background ............................................................. 1

MNH Register Field Trial ......................................................... 1
  Field Test Objectives ........................................................................ 2

Study Design ................................................................. 2
  MNH Data Collection Process Overview ........................................... 2
  Redesign of the MNH Register Form ................................................ 4
  Monthly Data Collection Process Specifics ........................................ 5
  HSA Training on New Form .......................................................... 7

Results ........................................................................... 8
  Form User Experience ................................................................. 8
  Field Observation ................................................................. 8
  Initial Form User Feedback .......................................................... 8
  Improving the Register Collection Process ....................................... 10
  Longitudinal Look at Form User Experience ....................................... 10
  Scanning in the field ................................................................. 11
  Register Durability ........................................................................ 12
  App User Training and Experience .................................................. 12
  Data Sync to Aggregate ............................................................ 15
  Number Recognition Accuracy ..................................................... 15
  Comparison to Manual Data Entry ................................................. 17

Conclusion ............................................................................. 18

Acknowledgements .................................................................. 19

Appendix A: Original MNH Register form ........................................ 20

Appendix B: ODK Scan-compatible MNH Register form ................... 21
ODK Scan Background
Developed in partnership with the University of Washington Department of Computer Science and Engineering, ODK Scan is an Android application that combines a device’s built-in camera with specialized image-processing technology to automatically digitize data from paper forms, simply by taking a photo. Optimized for use on a mobile device, the app quickly detects and processes hand-marked data, including bubbles, checkboxes and numbers, and translates the data into a digital format. As part of the Open Data Kit (ODK) suite of tools, ODK Scan is purposefully integrated with additional apps that allow ODK Scan users to edit, verify, aggregate, and create customized reports of collected information. Functional application of ODK Scan aims to significantly reduce the amount of time health workers and data managers spend on manual data input, allowing more valuable time and resources towards frontline health service and delivery. The ODK Scan project is funded by a $470,000 grant from the Bill & Melinda Gates Foundation to further develop and enhance the technology and support field testing.

MNH Register Field Trial
As one of the most resource-challenged countries in Sub-Saharan Africa\(^1\), Malawi is a prime candidate for streamlined data collection using inexpensive and portable information and communications technologies (ICTs). Furthermore, VillageReach has several ongoing projects in-country, many of which currently rely on paper processes for collection of data.

The program that was chosen for field testing ODK Scan is the Kwitanda Maternal and Neonatal Health (MNH) client program. In Malawi, Health Surveillance Assistants (HSAs) provide a variety of community-based health services, including conducting home visits during pregnancy. Data for each pregnant client is recorded on VillageReach-specific register books, which are then collected once a month by program officers. At the office, these client data are manually verified, digitized, and aggregated by the M&E Officer, who in turn creates monthly reports to update the funder and shape programmatic activities. share with VillageReach program leads and external donors. (See Figure 1.)

In this field test, the ODK Scan system was inserted into the data collection process in two main ways: Redesign of the MNH register using the ODK Scan form designer, and digitization of data using the ODK Scan app. The redesign of the register made scanning of the form with the Scan app possible, while testing the data digitization and review process gave important insights to the accuracy rates of the data classifiers. Furthermore, this field trial sought specific feedback from users on the usability of the redesigned form, ease of data field completion, app workflow intuitiveness, and overall experience with the new process.

Field Test Objectives

The objectives of this first field trial were to determine: 1) usability of the ODK Scan-compatible paper form by frontline health workers (HSAs); 2) data recognition accuracy; 3) usability of ODK Scan by field office administrative officers. Each of these were assessed as follows:

1) Usability of the paper form was assessed via interviews and surveys from the HSAs that participated in the field test. Specifically, we sought feedback on using the digit boxes to fill out numbers and the practicality/usability of bubbles and checkboxes for the remaining fields.

2) For data recognition accuracy, we analyzed accuracy of the number and bubble classifiers, and compared manual data entry accuracy versus automated digitization. In addition, we sought to determine if accuracy correlated to field snippet alignment, and how that could be used to strategize further improvement of the recognition classifiers.

3) For the ODK Scan app, we sought feedback on the user interface, intuitiveness of the inter-app workflow, the validation/editing process, the export process to ODK Aggregate. This included assessing the amount of time spent digitizing each register page and comparison with the previous data input process.

Study Design

MNH Data Collection Process Overview

The data collection process for this field test was designed to be analogous to the existing register data collection process where MNH data is collected monthly basis from the HSAs register books. In that process, seen in Figure 1 below, register books were collected between the 5th and 8th of every month and would remain at the VillageReach office in Balaka while the MNH data was entered into an Access database. During this time the HSAs would use alternate means of recording client visit information, as their register books remained inaccessible.
Even though ODK Scan is designed to make portable, remote scanning of forms possible, the field test plan was to maintain the practice of collecting MNH registers monthly, following the process shown in Figure 2. This was to enable the manual entry of the register data to continue in parallel, ensuring availability of data for programmatic needs and also enabling comparison of manually entered data vs. data entry via ODK Scan.
Based on initial feedback during the first two months of the trial, the process for the last month of field testing was changed to an in-field collection model. The parallel data entry for the third month was still maintained by referencing photos of the register pages. The reasons behind this process change are explored further in the “Improving the Register Collection Process in the Results” section.

**Redesign of the MNH Register Form**

In order to ensure both programmatic approval and technological capability of this field test, the first step was redesigning the MNH register form to be ODK Scan-compatible. This included assessment of which fields needed to be automatically digitized, if there were any fields that could be simplified, and how the original form’s field layout could be paralleled even with space constraints on the new form. Almost all of the number fields were turned into digit box fields on the ODK Scan-compatible form, while some free-response fields were converted to bubble fill-in fields containing common answers.

![Old form field vs. Redesigned ODK Scan form field](image)

**Date fields and other number fields were converted to digit boxes**

![Existing data showed that the “Location” response was typically one of three things, so the field was redesigned as a fill-in bubble field.](image)

Two QR code fields were also added to the page, one to help automatically identify the HSA responsible for that client and the other to provide automatic reference to the client’s record in the Access database. Both the original and redesigned forms can be found in their entirety in Appendices A and B.

After reviewing samples of printing and binding options for the register books, the team chose a plastic spiral binding style with one register form per page. This option (as opposed to cloth-bound) allowed the pages to sit flatter within the binding, enabling better scanning of the form. Once approved for use, these
redesigned register forms were printed into spiral-bound register books and prepared for distribution to the HSAs.

**Monthly Data Collection Process Specifics**

With the ODK Scan field trial, the process generally continued according to the existing schedule, with some trial-specific activities added. As noted in the process flow above, contracted data entry clerks input the data from the registers into the Access database. The M&E officer then scanned the same registers using the ODK Scan workflow and uploads the subsequent data to ODK Aggregate using the ODK Sync app on the tablet. Once this was complete, the books were returned to the HSAs in the field.

For the third month of data collection, the M&E officer traveled to the field and scanned the registers there, leaving the register books with the HSAs once the scans were complete. He then returned to the office and the data entry clerks were supplied with the full-page digital images of the registers (taken from the Scan photos) to use as references for data entry.

For the purposes of this 3-month study, we designed the scanning process to look at different processes for Scan:

<table>
<thead>
<tr>
<th>Month 1: June</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process:</strong> M&amp;E officer scans the registers in the Balaka office and ensures that the form has scanned correctly. He saves the form data without any editing or reviewing and uploads the data to ODK Aggregate using Sync.</td>
</tr>
<tr>
<td><strong>Research Objectives:</strong></td>
</tr>
<tr>
<td>• Establish ease of use and learnability of Scan app</td>
</tr>
<tr>
<td>• Measure time spent per page to take photo using Scan</td>
</tr>
<tr>
<td>• Assess ease of processed data synchronization with server</td>
</tr>
<tr>
<td><strong>App User Actions:</strong></td>
</tr>
<tr>
<td>1. Scan form &amp; save data <em>(ODK Scan)</em></td>
</tr>
<tr>
<td>2. Sync data with Aggregate server <em>(ODK Sync)</em></td>
</tr>
</tbody>
</table>
Month 2: July

Process: M&E officer scans the registers in the Balaka office and ensures the form has scanned correctly. He opens the processed data in ODK Survey and reviews six priority fields on the form. If any data in these fields has been processed incorrectly he edits the data and then saves the form. When the forms are complete, he uploads to ODK Aggregate using Sync.

Research Objectives:
• Establish ease of use and learnability of Scan to Survey app workflow
• Measure time spent per page to take photo and validate data
• Determine feasibility of data validation during collection process
• Assess ease of processed data synchronization with server over

App User Actions:
1. Scan form (ODK Scan)
2. Validate & save data (ODK Survey)
3. Sync data with Aggregate server (ODK Sync)

Month 3: August

Process: M&E officer travels to the field (Kwitanda district) and meets with HSA to collect their register data. M&E officer scans register page, and opens form in ODK Survey to review processed fields. Officer edits the data and then saves the form. When the forms are complete, he uploads to ODK Aggregate using Sync upon return to the Balaka office.

Research Objectives:
• Establish ease of use of Scan workflow in the field
• Measure time spent per page to take photo and validate data in field
• Assess challenges of in-field photo taking with Scan (shadows, locations, wind)
• Assess value add of in-person data collection to support programmatic goals

App User Actions:
1. Scan form (ODK Scan)
2. Validate & save data (ODK Survey)
3. Sync data with Aggregate server (ODK Sync)
**HSA Training on New Form**

Although the redesigned form made minimal changes to the existing MNH form, the HSAs needed to be trained on how to fill in the new number boxes and bubble fields. In order to do this, VillageReach held a one-day training where the new register books were introduced and the new field styles were explained. HSAs were given examples and guidelines for how to fill in the digit boxes using structured format numbers and were shown how to complete the fill-in bubbles. They were also asked to fill the fields in pen, rather than pencil. For practice, the HSAs were led through two sample client visits to give context to the new form instructions. Training leads worked one-on-one with each of the 18 HSA attendees to give individual feedback and answer specific questions.

At the close of the training HSAs were asked to take the new register books and begin using the new registers for all subsequent client visits.

The HSAs did not have any issues with learning how to fill the fill-in bubble fields correctly, and learning the specific number structures for each digit did not pose problems either. The main issues that arose centered on writing the actual numbers within the digit boxes. HSAs were instructed to use 0’s in dates and ID numbers, rather than leaving blank boxes. (For instance, an ID number of “5419” should be written
“05419,” filling each available digit box with a number.) Many of the HSAs struggled with this, especially since the variable lengths of clients’ ID numbers required an ID number field of five digit boxes on the form, even when many client IDs are only three or four digits. The challenge presented by necessitating writing these leading zeroes led the ODK Scan development team to prioritize and push an update to the app where empty boxes are better recognized as “blank.” This feature was implemented in time for the third month of data collection.

Results

Form User Experience

Field Observation
The day following the initial training, four HSAs were observed filling the new registers in the field. Each HSA took time filling the digits boxes, making sure to be careful to write the structured numbers in the prescribed format. Test scans of the forms using ODK Scan showed that many of the fields were easily and correctly recognized by the app. Of the numbers that were recognized incorrectly, many of those digit boxes had been filled out in the correct format and would be expected to process without a problem. This feedback was given to the HSAs and it was noted that even with seemingly perfect field filling that Scan would not process numbers with 100% accuracy (see Number Recognition Accuracy below).

Initial feedback from one HSA during these field visits was that he preferred the new register to the old, citing that he liked having the delineated digit boxes for writing numbers. In his opinion, it made it easier to see where to write the number and reminded him how many digits he should be recording. He also liked the guide dots within the digit boxes because it made his numbers appear neater. The other HSAs did not volunteer any information about a preference between the new and the old registers, but were open to using the new registers for subsequent client visits.

Initial Form User Feedback
The field test team conducted two subsequent sessions with the HSAs in order to gather usability feedback on form user experience over time and solicit suggestions for both form and process improvement. The first session took place after the HSAs had been
using the form for two months, and the second session took place after five months total. Feedback was solicited using three different activities: a strengths and challenges Post-it Note icebreaker, an anonymous survey, and one-on-one interviews. The majority of the questions on the survey were Likert scale assessments where HSAs were asked to rate the ease/difficulty of different aspects of using the forms. The one-on-one interviews were conducted in English and Chichewa (as needed) by program organizers and were guided by predetermined question topics. See Appendix D for full surveys and interview questions.

The first feedback session was intended to give preliminary feedback on form user experience and inform any process changes that may have been needed for the remainder of the field test. The second was to capture a more longitudinal understanding of how form ease of use may have increased or decreased as the HSAs continued to use their new registers. It also gave a chance to capture information on the physical durability of the new register books in comparison to the older version.

While the majority of HSAs reported that learning to fill the form was not overly challenging, actually writing data on the form was not as straightforward. Filling of bubble fields was generally reported as being very easy (72% of respondents), whereas filling of the number fields was much more difficult, with 50% of respondents reporting it “not very easy,” and 17% indicating the lowest marker at “not easy at all (difficult). HSAs reported that completing the number fields in the method instructed with the digit boxes was time consuming and difficult. Two HSAs specifically called out their frustration that when they make a mistake on the form they are not able to correct it, as they had been instructed to use pens to complete the number fields. This, they indicated, caused them to be discouraged and made them less inclined to fill the rest of the form. When asked for suggestions to improving the form, several HSAs requested that the forms be changed so that the old method for writing numbers was reinstated—that is, writing free-form numbers without digit boxes or prescribed structures. Other HSAs suggested that they be able to use pencils to fill the forms, as that way any errors could be easily corrected. Full data from the surveys is in Appendix C.

Alternatively, some HSAs asked that number fields be converted into the fill-in bubble fields. This type of field was very popular for the HSAs, and they rated its ease of use accordingly high. In the one-on-one interviews, HSAs noted that these bubble fields were convenient not only because they were simple and quick to fill, but because in the redesigned form fill-in bubble fields replaced five of the number or free

---

2 Although the data collection for this field test concluded at month three, the HSAs continued to use the redesigned registers to collect MNH information, and as of December 2015 are still using the ODK Scan-compatible registers.
response fields. This perceived improvement in form design may have been a reason that so many HSAs reported that they prefer the new register form over the old one, even with the challenge of the structured number fields.

**Improving the Register Collection Process**

Outside of feedback on the register form itself, HSAs were asked if they had suggestions for overall data collection process improvement. The most commonly cited issue was frustration with the registers being inaccessible for two weeks each month and the extra work that this unavailability causes for the HSAs. When the registers are collected for monthly data collection, the HSAs must find an alternative method for recording client visit information. The HSAs continue to visit clients while the registers are at the office, but they must record the visit information on loose sheets of paper, sometimes paid for out of their own pocket. When the registers are returned to the field, the HSAs must then spend extra time transferring the interim visit data into the register books. As expected, spending extra time on duplicated efforts is not an efficient or popular process. While this issue predates the ODK Scan field trial, it was recognized that the mobile nature of data collection using the app could help resolve this inefficiency.

Based on feedback from the form and app users, the ODK Scan team worked with MNH program managers to pursue a modified register data collection process for the remainder of the field test. Instead of collecting the register books from the HSAs and taking them to the VillageReach office for data entry, it was decided that the data collection for month three would take place in the field with the HSA present. While in the field, the M&E officer would use the same process of scanning and validating the form, but would then return the register to the HSA after completing the register scans. He would then sync all data with the Aggregate server upon returning to the office.

This change in process further evidences the need for a portable data collection system that does not remove the paper record from the health worker. Even without experiencing direct benefit from automatic digitization of fields, the HSA still experiences a huge benefit from retaining access to their essential paper system. This is a major advantage that ODK Scan can offer to health systems at the last mile, as the Scan workflow enables improvement of even some non-digital processes.

**Longitudinal Look at Form User Experience**

The second feedback session with the HSAs took place five months after the initial training and focused on eliciting information on form use over time and experiences with the registers being scanned in the field. Survey results show that the average ease of use in navigating the form and filling bubble fields remained the same. Ease of use for filling digit fields showed the biggest change with the average assessment increasing two degrees, from “not very easy” to “somewhat easy.” This was further supported by responses from the individual interviews, where nearly 70% of HSAs reported that the digit boxes became easier to fill over time.
Even with this improvement, however, it is clear from the interview responses that the HSAs would prefer to return to the old system of writing numbers, as it does not pose the same challenges with number structure, fixing errors, and time spend. These findings are key considerations in future form redesign, and have encouraged the ODK Scan team to focus on implementing digit box fields only for scanned data that is used at a programmatic level. For example, an HSA needs to be able to verify that a woman has visited her antenatal care (ANC) clinic, and for that they look at the date in the woman’s health passport and record it in the MNH register. However, this specific date information is not used at an aggregated program level. Rather, it is simply the successful completion of the ANC visit that is necessary to inform the aggregated monthly reports. As such, the ANC visit field could be converted from a number box to a freehand text box combined with a bubble “fill if yes” field. This would minimize the number of fields requiring the more challenging number box format but still allow automatic digitization of data needed at the program level.

**Scanning in the field**

When asked about the in-field register scanning as compared to the register collection and scanning at the office, HSAs overwhelmingly (95%) responded that they preferred the in-field process. Many respondents noted that the in-field process was less disruptive to their work and that they benefitted from having access to their registers for the whole month. Many also noted that they valued the monthly touch point opportunity with the M&E officer, as they were able to ask questions and/or clarify information related to their work.

Overall, the HSAs reported an good experience with using the registers (Figure 4), and especially with the new in-field collection process. Future program efforts are already planned to integrate the benefits of the new collection process with an improved form user experience, streamlining timely and quality data flow from the last mile.
Register Durability
Collecting longitudinal feedback from HSAs also allowed an opportunity to assess how the physical registers held up over time. Unlike the old registers, which were cloth-bound with a cardboard cover, the new registers were spiral-bound with a plastic cover. HSAs noted that the spiral binding caused the pages to rip out more easily, which was not desirable. However, they also noted that the new plastic covers were much more durable than the previous cardboard ones, and requested that future registers include a hybrid of the cloth binding and plastic cover page. Additional prototyping and testing would be needed to determine if fabric binding would allow pages to lay flat enough for scanning, and will be investigated further in any subsequent field work.

App User Training and Experience
The M&E Officer was introduced to the ODK Scan project during the field trial planning process and was trained one-on-one on the app during the first Scan deployment trip in Malawi. He was provided with an Asus Nexus 7 tablet pre-loaded with the Scan, Survey, Tables, and Sync apps and was instructed on how to move files between applications. Most importantly, the M&E officer was trained on how to scan the register pages using the tablet, making note of recommendations regarding how to hold the device, optimal lighting, and form alignment. Because the scanning process for this field test takes place in an office setting, the M&E officer was also provided with a plastic stand to place the device on for easier photo taking. The M&E officer used this stand during all three months of data collection, including scans done in the field.
While training took place in person, follow-up support provided remotely over email and Skype. Overall, the process was much less intensive than expected, as the M&E officer showed he had a high comfort level with the tablet technology and was able to learn the basic Scan workflow within the first hour of training.

As noted in the collection process section, data entry using Scan varied between the first, second, and third months of collection. In month one, the M&E officer scanned each page of the register book, confirmed that the form had scanned successfully, and saved the record. When timed, this process took the M&E officer approximately 20-30 seconds. After all the records were scanned successfully and saved, the officer synced the records with Aggregate, without first viewing or editing any individual data points. With a total of 200 records to send to ODK Aggregate, this syncing process took over two hours on the office Wi-Fi connection.

For the second month, the M&E officer scanned each register page and then also reviewed 3-6 pre-selected fields, depending on whether the client was a new registrant or an existing client. The M&E officer was asked to manually input the client’s name and then validate the automatically processed fields. Once all validation was complete, he saved the record and then uploaded the completed records, syncing all the records in bulk at the end of data collection. The process of scanning and validating the form fields took approximately two minutes per page. This was a significant improvement from early-stage field testing in 2012 in Mozambique. In those results, users took an average of 12 minutes to scan and validate each form. This significant reduction of time spent on each page can be attributed both to technical improvement of the in-app Scan processing and to prioritized listing of relevant fields within ODK Survey validation. The inclusion of a “table of contents” style menu in Survey allowed the user to tap into specific fields for editing, rather than needing to swipe through all fields to get to the desired one.

After validation of the data was complete for all forms in that collection (about 200 per month), the M&E officer used ODK Sync to conduct a bulk synchronization with the ODK Aggregate cloud database. This sync over the office Wi-Fi took over two hours, highlighting the need for further network and file size optimization. This is further addressed in the Sync Optimization Efforts section below.

---

3 Dell, N., Crawford, J., Breit, N., Chaluco, T., Coelho, A., McCord, J., Borriello, G. (2013). Integrating ODK Scan into the community health worker supply chain in Mozambique. ICTD 2013, Cape Town, South Africa: ACM.
In exploring ODK Scan app usability, the M&E Officer reported that Scan was very straightforward and easy to use, and he had no suggestions for necessary improvements. He did not experience any bugs with Scan or the other ODK apps, and over the two first months of the field test has achieved fluency with the app workflow. He did note, however, that the tablet is not optimized for use in the office and that alternative devices and practices (such as utilizing an external keyboard) might be needed for strictly office-based scan and validation. These observations helped drive the process change for the third month of data collection, where in-field register scanning was trialed. It was discussed that in recreating the old MNH register data collection process for this field test, the Scan app was not providing the optimum benefit that it could be. ODK Scan was conceived and developed as a mobile tool, designed with specific functionality to work in the field and outside of a traditional office setting, and there was untapped potential to both improve the register process and test the mobile aspects of Scan. The M&E officer suggested that another real benefit to data collection in the field would be the opportunity for immediate feedback to the HSAs on their register data. He noted that sometimes an HSA misses fields that need to be completed, and by utilizing in-field data collection he could address problems and have the HSA make any register changes immediately. These scanning sessions took place at the Kwitanda clinic, the antenatal clinic, or, in two cases, at the home of the HSA.

In giving feedback after the third collection, the M&E officer reported that although the scanning took 30 seconds longer per page than in the office (due to increased distractions, questions, etc.), there were added benefits to the in-field scanning. In addition to providing a valuable point of contact resource for the HSAs, the M&E officer was also able to use the one-on-one time to review specific fields on the register that had missing or unclear data. In previous months, he would need to call the HSA for the information, but having the HSA present at the time of scanning meant that he was able to resolve things at that very moment. The M&E officer noted that while these trips to the field limited the other work activities he could complete during the day, he emphasized that these field visits offered more value than simply allowing the HSAs to retain their registers for the full month, and could be an asset to future programmatic activities. His main suggestion for field scanning was that in-field validation using ODK

Main User Feedback on App

- App workflow is easy to learn
- ODK Scan itself is straightforward and did not present challenges
- Tablet format is not optimized for data editing and validation can be time-consuming
- Suggestion for bulk data validation rather than page by page (does this contradict earlier statement that no improvements were suggested?)
- Long sync times and potential for errors is problematic
- In-field scanning offers process and programmatic value

Image 5: Scanning registers in the field at the Kwitanda ANC clinic meeting area
Survey took time and that validation of fields after returning to the office would be preferred.

Data Sync to Aggregate
Although scanning and validating data did not pose any major challenges during the field trial, the process of syncing the data with the Aggregate server proved to be problematic. While the ODK Sync app was reported as easy to use, there were issues with the sync process taking much longer than expected, sometimes taking hours to complete data upload. This proved to be the only major pain point in using the app workflow and highlighted the need for focused software development and configuration on optimizing sync time.

For this particular register form, the data count comes in around 550KB per scanned page, with about 85% of that data coming from the image snippet files. At 200 pages scanned per month, each month’s sync uploaded around 110MB of data to the Aggregate server. A variety of different issues caused delays and errors with the sync process, but the ODK Scan team has already begun investigating how to optimize this upload post-trial. Switching from Google AppEngine to and AWS server has already cut syncing time in half, while changing the check-in process with Aggregate could yield more efficient syncing with the device connection. Based on findings from the trial, the team has changed the sync process to instead of uploading each image individually, the system is does batch uploads per row. This, along with additional Sync work prioritized on the roadmap, will continue to refine the data upload process to the server.

Number Recognition Accuracy

With over 600 MNH register forms scanned at 50+ fields per page, the ODK Scan field test in Malawi has gathered a large and useful data set from whence to calculate real-world data recognition accuracy rates. These data have been filtered and analyzed through careful visual manual verification of selected fields using custom data export spreadsheets built by the ODK Scan development team. Because there is such a large amount of data, the team opted to analyze a portion of the total fields scanned. The team randomly selected 100-150 pages from each month’s collection and manually verified 15 fields from each page. These fields were the same across all pages and were preselected to reflect diverse field styles (basic number, vertical bubble, date, etc.) and locations on the form. This resulted in analysis of approximately 6,000 fields made up of over 16,000 individual numbers and 4,500 individual bubbles. By looking at the image snippet of each field and comparing it to the value that Scan processed, the calculated accuracy rates were determined as follows:

<table>
<thead>
<tr>
<th>Field Type</th>
<th>Month 1 (office)</th>
<th>Month 2 (office)</th>
<th>Month 3 (in field)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>85.43%</td>
<td>84.15%</td>
<td>80.25%</td>
<td>83.28%</td>
</tr>
<tr>
<td>Fill-in Bubble</td>
<td>96.80%</td>
<td>97.94%</td>
<td>92.92%</td>
<td>95.89%</td>
</tr>
<tr>
<td>Combined</td>
<td>91.12%</td>
<td>91.05%</td>
<td>86.59%</td>
<td>89.59%</td>
</tr>
</tbody>
</table>

Accuracy Rate by Month

For this data analysis, fields written in pencil and those with completely blank answers were thrown out.
As seen in the data above, the data accuracy rates for numbers and stay somewhat consistent for the first two months and then decrease for the in-field collection during the third month. This is somewhat expected, as the field scanning did not provide for optimal lighting and scanning stand surface conditions. An increased amount of shadows on the page or additional bends in the paper can have a significant effect on the accuracy classifier.

In comparison to limited lab testing, where numbers were scanned at 95% and bubbles were scanned at 99%, these field trial rates are also come in low. While less-than-optimum rates are expected outside of the lab, the ODK Scan team wanted to investigate the possible reasons for this discrepancy and identify what next steps should be taken to improve data recognition accuracy. One possibility for low rates could have been that the HSAs were not writing the numbers in the structured format, but analysis of sampled images from scanned forms indicates that the HSAs are adhering very closely to the prescribed format, with many of the numbers even being “textbook” examples. Another possibility is that the app user was capturing blurry or poor-quality photos during the scan, but analysis of scanned images also shows that that is not the case. The ODK Scan team then posited that the most likely reason for the decrease in accuracy rates is improper alignment of the data fields within the processed form. No matter how powerful or accurate a classifier is and how perfect the written numbers are, if the application is not looking in the right place on the form for the number, the data will be erroneous.

To investigate this hypothesis, each field analyzed was also assessed for alignment within its image snippet and accuracy rates were recalculated. The results (averaged across months) are as follows:

**Accuracy Rate by Field Alignment**

<table>
<thead>
<tr>
<th>Misalignment factor</th>
<th>No Misalignment</th>
<th>Small Misalignment</th>
<th>Medium Misalignment</th>
<th>Large Misalignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers</td>
<td>92.87%</td>
<td>88.25%</td>
<td>87.72%</td>
<td>68.79%</td>
</tr>
<tr>
<td>Fill-in Bubbles</td>
<td>98.75%</td>
<td>98.34%</td>
<td>97.04%</td>
<td>85.08%</td>
</tr>
<tr>
<td>Overall</td>
<td>95.81</td>
<td>93.29</td>
<td>92.38</td>
<td>76.94</td>
</tr>
</tbody>
</table>

---

This misalignment within the app is likely due to sub-optimal lighting in the photo and skewing of form fields due to bends in the page. In the case of extreme issues in the photo, Scan requires a retake from the user, but further honing of the alignment classifier to refine what is acceptable and with is not can improve the alignment of the images Scan does process.
It’s that that for both number and fill in bubble fields, recognition accuracy decreases as misalignment of the field increases. It would then be fair to argue that if overall alignment of the fields increased, the accuracy rate would as well. This line of thinking led the ODK Scan team to identify optimizing alignment algorithms in the Scan software as the next step in improving data recognition. If pages were successfully scanned and processed with no misalignment for their fields, average accuracy rates for numbers collected in the field could subsequently increase from 83% to 93%. This would be a significant gain for Scan’s recognition accuracy, and the ODK Scan development team has already identified an alignment improvement plan for the next phase of software optimization.

Comparison to Manual Data Entry
Knowing that ODK Scan’s recognition accuracy will never be 100% perfect, it became imperative to understand the maximum data accuracy rates for the most popular digitization alternative: manual data entry. Since the data collection process for this trial also included a parallel manual data entry process, comparison of the image snippet and Access database data was conducted, resulting in some rough manual data accuracy rates. From a sample of 500 fields, comparing between “ground truth” data from field image snippets and register data manually input into the Access database, the overall manual data entry accuracy rate comes in at 94%. In this analysis, administrative edits commonly made to certain fields were taken into account and not counted as incorrect. Also, fields that were left blank in the Access database that had a corresponding value on the form page were thrown out of this calculation. When these blank Access fields are included, the accuracy rate for manual entry drops to 88%.

ODK Scan vs. Manual Entry Accuracy Rates

<table>
<thead>
<tr>
<th></th>
<th>Overall Scan</th>
<th>Overall Manual</th>
<th>Scan with no misalignment</th>
<th>Manual with incorrect blanks included</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>91.92%</td>
<td>94%</td>
<td>96%</td>
<td>88%</td>
</tr>
</tbody>
</table>

As expected, the manual data entry rate is also not perfect, and an increase to 93% accuracy in the ODK Scan recognition classifier would actually yield comparable accuracy results to manual data entry in this context offers. It is also essential to note that the ODK Scan accuracy rate is calculated before any data validation takes place, whereas the Access values have already been reviewed.
As this is an analysis of a sampling of manually entered data, a more robust analysis (double-blind study) of Access and image snippet data is required to make a definitive conclusion about manual data accuracy. Nevertheless, this data serves as an important reminder that technology limitations are but one resource constraint of health systems at the last mile.

**Conclusion**

This field trial provided an essential first look at how the ODK Scan alpha software and integrated apps can perform in an actual last mile setting. User feedback and field observation asserted the feasibility of using ODK Scan compatible forms in place of existing paper forms and provided essential findings for how to smooth the transition. Understandings of the number field usability challenges and strong preferences for bubble fields has already been incorporated into form redesigns for future trials. Furthermore, the unanticipated value add of field-based register scanning has further emphasized the advantage of ODK Scan as a mobile solution. As many of the HSAs had previous experience with direct-to-digital mobile apps, they also provided the assessment that they preferred using a phone to collect data, but it was rare that both the app and the phone were working properly. Due to this, they noted that a paper register solution was the most reliable for their work. Anecdotally, this supports the advantage of a tool like Scan that maintains a consistent paper system.

Future field trials will focus on extended data analysis of time gains in data collection and district-level reporting due to Scan process implementations. This will include time spent on manual filling of existing forms compared to Scan forms and higher-level analysis of how Scan can improve timely reporting of aggregated data. It is essential in the next trial to quantify how useful Scan at a programmatic level, ideally evidencing that Scan-collected data can be easily integrated into existing program databases and systems.

As investigated in the data analysis from this field trial, improvements to the alignment classifier could significantly increase the current accuracy rates. With a predicted maximum accuracy of 93% for numbers, this rate falls slightly under our grant goal of 95% accuracy. However, preliminary analysis of manual data entry rates show 93% to be nearly equal to the accuracy of manual data entry (94%), and arguably already more resource-efficient. This alignment work, combined with the sync optimization efforts described earlier, form the core of software development efforts for Scan over the next six months. Improvements to ODK Scan workflow and form usability will also be integrated into future work, striving to find an impactful balance between human-driven data collection and technology-enabled data processing. Overall, this field trial has helped to assert the strengths and potential impact of ODK Scan, refining the opportunity for impact that this special tool holds for last mile health systems.
Acknowledgements

Core ODK Scan team:
Rachel Powers – VillageReach
Sarah Jackson – VillageReach
Jeffrey Beorse - University of Washington CSE
Richard Anderson – University of Washington CSE
Nicola Dell - University of Washington CSE

The ODK Scan team would like to acknowledge the following groups and individuals for their involvement and essential support of this ODK Scan field test in Malawi:

Robert Saiti, for his key and irreplaceable role in project organization, field trial activities, and app testing;
Jodi-Ann Burey, Taiki Endo, Erin Larsen-Cooper, Jessica Crawford, Barbara Jones Singer, Beatrice Chaponda, Christopher Feza, and Saiti Chikwapulo, for their essential organizational support;
Clarice Larson, Mitch Sundt, and Waylon Brunette, for their Open Data Kit technical support;
And for their crucial involvement and patience, the Kwitanda District Health Surveillance Assistants: Acklen Mdoka, Asnty Chikapaza, Beatrice Chikaoneka, Ben Banda, Catherine Kazembe, Charles Isaac, Chisomo Chisada, Cosmas Bakili, Emma Muwa, George Chidulo, George Koloko, Gladys Kaselo, Luka M'bawa, Mabvuto Mandere, Mackson Khalawako, Maxwel Kadzilawa, Mercy Macholowe, Pemphero Katondo, Rodney Kamaliza, and Tobias Kadzoma.

Lastly, this work is a continuation of the vision of Gaetano Borriello to develop mobile tools to strengthen health systems. Gaetano passed away prematurely in February, 2015. We are grateful for his wide-ranging contributions to humanitarian technology and to the development of impactful solutions at the last mile.

Thank you.
Appendix A: Original MNH Register form

VillageReach Maternal and Newborn Health Supplemental Register for HSA:

<table>
<thead>
<tr>
<th>Name:</th>
<th>ANTE NATAL CARE</th>
<th>Maternal Health Conditions (tick if yes)</th>
<th>HIV Testing (tick if yes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(last name, first name)</td>
<td>ANC Location:</td>
<td>Hypertension/Pre-eclampsia</td>
<td>[ ]</td>
</tr>
<tr>
<td></td>
<td>Month (#) of pregnancy when ANC started:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID #:</td>
<td>Dates of ANC Visits</td>
<td>Tetanus Toxoid Vaccine</td>
<td>Diabetes</td>
</tr>
<tr>
<td></td>
<td>#1</td>
<td>#1</td>
<td>Under the age of 20</td>
</tr>
<tr>
<td>Age:</td>
<td>Village:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#2</td>
<td>#2</td>
<td>Underweight</td>
</tr>
<tr>
<td></td>
<td>#3</td>
<td>#3</td>
<td>Carrying twins or triplets</td>
</tr>
<tr>
<td>Maternal History</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#4</td>
<td>#4</td>
<td>History of preterm delivery</td>
<td>[ ]</td>
</tr>
<tr>
<td>EDD:</td>
<td></td>
<td></td>
<td>Excessive vaginal bleeding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total # of pregnancies:</td>
<td>Malaria Prophylaxis</td>
<td>Iron Given (tick if yes): [ ]</td>
<td>death</td>
</tr>
<tr>
<td>Total # of live births:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1</td>
<td>Albendazole</td>
<td>Other:</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>#1</td>
<td>Other:</td>
<td></td>
</tr>
<tr>
<td>CCPF</td>
<td>ITN Given (tick if yes): [ ]</td>
<td></td>
<td>Severe Headache</td>
</tr>
<tr>
<td>Registered for CCPF: Yes No</td>
<td>POSTNATAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>If no, registration form completed?</td>
<td>Date of Delivery:</td>
<td>Postnatal Checkup (tick if yes): [ ]</td>
<td></td>
</tr>
<tr>
<td>Yes No</td>
<td>Location of Delivery:</td>
<td>Date Attended:</td>
<td></td>
</tr>
<tr>
<td>Complications/Notes:</td>
<td>Location:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Official Use Only:</td>
<td>Infant Status</td>
<td>Vitamin A given: [ ]</td>
<td>Breastfeeding: [ ]</td>
</tr>
<tr>
<td>Registration Complete:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharge Complete:</td>
<td>Live Birth ____ Stillbirth ____ Neonatal Death ____</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complications/Notes:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Home Visits and Education (Place a check mark next to each topic)

<table>
<thead>
<tr>
<th>Visit #</th>
<th>Date</th>
<th>Pregnancy Danger Signs</th>
<th>Malaria Prophylaxis</th>
<th>HIV/TB Counseling</th>
<th>Activity Level</th>
<th>Tetanus</th>
<th>Nutrition</th>
<th>Birth Plan</th>
<th>Breastfeeding</th>
<th>Family Planning</th>
<th>Postnatal Danger Signs</th>
<th>Neonatal Care/Danger Signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: ODK Scan-compatible MNH Register form

Malaria Prophylaxis given (fill if yes):
#1
#2
#3
#4
#5

Albendazole given (fill if yes):
#1
#2

Iron given (fill if yes):

ITN given (fill if yes):

HIV Testing (fill if yes):

HIV Testing (fill if yes):

HIV Testing (fill if yes):

Birth Plan:

Location:

Transport (select one):
- bicycle
- walk
- car/bus
- other

Danger Signs (fill if yes):
- Excessive vaginal bleeding
- Fever
- Foul vaginal discharge
- Convulsions
- Severe Headache

ADDITIONAL NOTES (include information about miscarriages, abnormal findings, or any other relevant details.)

Vitamin A given:
Yes
No

Breastfeeding:
Yes
No

Complications/notes:

Live Birth
Stillbirth
Neonatal Death

Pregnancy Danger Signs
Malaria Prophylaxis
HIV/TB Counseling
Activity Level
Nutrition
Birth Plan
Breastfeeding
Family Planning
Postnatal Danger Signs
Neonatal Care/Danger Signs

Visit #

#1
#2
#3
#4
#5

Date

Appendix B: ODK Scan-compatible MNH Register form
## Appendix C: Quantitative Feedback from HSA Feedback Sessions

<table>
<thead>
<tr>
<th>Question</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How has your overall experience been with the new register?</td>
<td>Very good</td>
</tr>
<tr>
<td>Feedback Session #1 (at 2 months)</td>
<td>36.84%</td>
</tr>
<tr>
<td>Feedback Session #2 (at 5 months)</td>
<td>33.33%</td>
</tr>
<tr>
<td>Percent Change</td>
<td>3.51%</td>
</tr>
<tr>
<td>Overall Assessment</td>
<td>Overall experience increased.</td>
</tr>
<tr>
<td></td>
<td>Overall ease of learning to fill out register increased. Very easy the most common response, changing from the top response of Not very easy.</td>
</tr>
<tr>
<td></td>
<td>Overall ease of navigating form fields decreased slightly.</td>
</tr>
<tr>
<td></td>
<td>Overall ease of writing numbers increased from Not very easy (50% @ first feedback meeting) to somewhat easy (42% @ second meeting).</td>
</tr>
<tr>
<td></td>
<td>Overall ease of filling out bubbles still reported as Very Easy, but percentage decreased from 72% to 57%. All except for one remaining response reported Somewhat Easy.</td>
</tr>
<tr>
<td></td>
<td>Overall opinion on comparative time to fill new form still reported as more time than old form.</td>
</tr>
</tbody>
</table>
Appendix D: Survey and Interview Questions

1. How has your overall experience been with the new register?

☐  

Very good  Somewhat good  Neutral  Not very good  Not at all good (bad)

2. In your opinion, how easy was it to learn how to fill out the new register?

Very easy  Somewhat easy  Neutral  Not very easy  Not at all easy (difficult)

3. In your opinion, how easy is it to find the fields you need on the register page? (Is the layout easy to navigate?)

Very easy  Somewhat easy  Neutral  Not very easy  Not at all easy (difficult)

4. In your opinion, how easy is it to write the numbers using the guide dots in the number boxes?

Very easy  Somewhat easy  Neutral  Not very easy  Not at all easy (difficult)

5. In your opinion, how easy is it to fill the bubbles for the bubble fields?

Very easy  Somewhat easy  Neutral  Not very easy  Not at all easy (difficult)

6. In your opinion, does the new register take more time, less time, or equal time to fill out compared to the old register?

More time  Less time  Equal time

7. Is there anything you don’t like about the new register?
8. Is there anything you do like about the new register?

9. What improvements would you suggest to make the register easier to use?
HSA On-on-one interview questions:

1. Describe your experience with the new register.

2. Describe your experience when you first started writing the numbers in the digit boxes.

3. Describe your experience with writing the numbers in the digit boxes now.

4. Did filling in the numbers using the number boxes get easier, harder, or did it stay the same?

5. What has been your experience with the durability of the register?

6. How does it compare to the durability of the old register? (Are the issues the same, or are there new issues?)

7. Describe what happened when your register was collected for data entry.

8. Describe how you recorded MNH information while your register was at the VillageReach office. (Where did you write the information? How did you acquire additional stationery?)

9. Describe what happened when your register was scanned and validated in the field.

10. Describe your experience with Robert taking pictures of your register in the field.

11. What did you like about the register scanning in the field in August?

12. What did you dislike about the register scanning in the field in August?

13. Do you prefer handing your registers for monthly collection or having them scanned in the field? Why?

14. How did the process of scanning the register in the field affect your job? (Did it make it easier, harder, or did nothing change?)
15. Do you have any additional feedback about the registers or data collection process?