

Article

ODK Scan: Digitizing Data Collection and Impacting Data Management Processes in Pakistan's Tuberculosis Control Program

Syed Mustafa Ali ^{1,*}, Rachel Powers ², Jeffrey Beorse ^{3,*}, Arif Noor ¹, Farah Naureen ^{1,*}, Naveed Anjum ¹, Muhammad Ishaq ¹, Javariya Aamir ¹ and Richard Anderson ³

¹ mHealth Coordinator, Mercy Corps, Pak Palace, Murree Road, Rawal Chowk, Islamabad 44000, Pakistan; anoor@mercycorps.org (A.N.); nbuzdar@mercycorps.org (N.A.); muishaq@mercycorps.org (M.I.); jaamir@mercycorps.org (J.A.)

² Information Systems, VillageReach, Seattle, WC 98102, USA; rachel.powers@villagereach.org

³ Computer Science and Engineering, University of Washington, Seattle, WA 98101, USA; anderson@cs.washington.edu

* Correspondence: symustafa@mercycorps.org or mustafa_30_84@yahoo.com (S.M.A.); jbeorse@cs.washington.edu (J.B.); fnaureen@mercycorps.org (F.N.); Tel.: +92-30-1855-0454 (S.M.A.); +1-206-221-7701 (J.B.); +92-30-2850-6178 (F.N.)

Academic Editor: Dino Giuli

Received: 30 August 2016; Accepted: 12 October 2016; Published: 24 October 2016

Abstract: The present grievous tuberculosis situation can be improved by efficient case management and timely follow-up evaluations. With the advent of digital technology, this can be achieved through quick summarization of the patient-centric data. The aim of our study was to assess the effectiveness of the ODK Scan paper-to-digital system during a testing period of three months. A sequential, explanatory mixed-method research approach was employed to elucidate technology use. Training, smartphones, the application and 3G-enabled SIMs were provided to the four field workers. At the beginning, baseline measures of the data management aspects were recorded and compared with endline measures to determine the impact of ODK Scan. Additionally, at the end of the study, users' feedback was collected regarding app usability, user interface design and workflow changes. A total of 122 patients' records were retrieved from the server and analysed in terms of quality. It was found that ODK Scan recognized 99.2% of multiple choice fill-in bubble responses and 79.4% of numerical digit responses correctly. However, the overall quality of the digital data was decreased in comparison to manually entered data. Using ODK Scan, a significant time reduction is observed in data aggregation and data transfer activities, but data verification and form-filling activities took more time. Interviews revealed that field workers saw value in using ODK Scan, but they were more concerned about the time-consuming aspects of the use of ODK Scan. Therefore, it is concluded that minimal disturbance in the existing workflow, continuous feedback and value additions are the important considerations for the implementing organization to ensure technology adoption and workflow improvements.

Keywords: mHealth; ODK scan; mobile health application; digitizing data collection; data management processes; paper-to-digital system; technology-assisted data management; treatment adherence

1. Introduction

Tuberculosis (TB) continues to be a global and grave public health concern. It is a health issue for millions and the cause of numerous deaths each year worldwide [1]. Despite the availability of effective treatments and strategies to improve access to treatment regimens, multidrug-resistant and

extremely drug-resistant TB cases are emerging due to the lack of adherence to treatment protocols. This non-adherence is mainly because of the longer treatment course (6 to 8 months) and unpleasant side effects of the drug treatment [2].

Pakistan is a TB high-burden country, experiencing the emergence of a large number of new TB cases (approximately 420,000 every year) and prevalence of multidrug-resistant TB. Pakistan ranks fifth for drug-susceptible TB and fourth for multidrug-resistant TB [3]. The burden of disease caused by TB and the effectiveness of different programmatic interventions are better understood through an interpretation of the available epidemiological data [4]. The program management function, Monitoring and Evaluation (M&E), makes the best use of the available resources (including data) to guide and support programmatic decision-making. The M&E system plans and conducts data collection and analysis to inform program managers and keep performance measures on track [5]. Unfortunately, this immense resource is either underutilized or not used at all [4].

Health information systems used in disease-specific health programs are designed to collect patient-centric data to generate actionable information to guide decision-making. Likewise, TB control programs mostly maintain and manage paper-based patient recording and reporting systems [6].

Fidelity to treatment protocols is ensured through case management (DOTS strategy) and follow-up evaluations (sputum specimen for microscopic examination or culture at specific points in the treatment period) [7]. Generally, in the TB control programs, paper-based patient records are used to prepare periodic reports on the patient enrolment, retention and treatment progress. These data summaries are also sent to different management levels for aggregation, analysis and reporting [6]. However, paper-based recording and reporting systems suffer from some inherent challenges, which include: checking data quality, data entry and aggregation, allotting treatment outcome correctly, filling gaps in data, analysing data and generating actionable information. These data management activities are inefficient, work-intensive and time consuming [8]. However, the advent of digital technology has positively impacted the organizational workflow [9].

Electronic Health Information Systems (eHIS) have generated better results on the three fronts of TB control programs: patient care, resource management and disease surveillance [8]. Increasingly, mobile health (mHealth) interventions are grabbing the attention of the public health professionals because of their ability to improve the public health situation [10]. mHealth refers to the application of portable hand-held devices and communication capabilities in the health care programs [11]. During the past decade, mHealth has been continuously evolving within the eHealth domain, where it has shown great potential to overcome many traditional barriers in delivering care and reducing the associated distance, time and cost [12]. Fortunately, most developing countries have already launched mHealth initiatives to unleash the power and reach of smartphones and consider them an important resource for front line workers [10].

Early diagnosis and vigilant management are important considerations to successfully control TB. Therefore, in TB control programs, case management involves a multidisciplinary approach for comprehensive and coordinated follow up of the presumptive and confirmed TB case [13]. This approach includes improved communication between patient and care giving teams, patient education and delivery of support during treatment [14]. Increasingly, mHealth applications are being used to support health care teams in early diagnosis and monitoring patient progress [15]. Many of the existing mHealth solutions for TB case management are custom software developed for individual organizations and use cases, they are not freely available for other organizations to try. There are also general purpose mHealth solutions available, such as Medic Mobile or ODK Collect and its cousins. These are direct-to-digital systems that do not involve any paper record keeping or paper archives.

Among the various reported uses of mHealth tools the most reported are: data collection, electronic decision support, provider's education and training, and communication [10]. However, the particular use of mHealth for data collection has a greater likelihood of improving public health in low- and middle-income countries [16]. Noteworthy are those unpredictable implementation

challenges and complicated situational factors that affect an introduction of mHealth applications in primary healthcare settings of the developing world [17,18].

Since there are a number of freely-available open source mHealth applications aimed at improving data collection, only one tool leveraged the strengths of the existing paper system with the flexibility and efficiency of the digital records. We chose the ODK Scan paper-to-digital application to test improved data collection in the public-private mix model of the TB control program of Pakistan. The ultimate goal of this intervention was to reduce the time and resources spent on TB data collection by improving the data collection and digitization process and reducing reliance on manual data entry. Additionally, ODK Scan was chosen because it allows supervisors and managers to assess the data quality by reviewing the original image snippets. The specific objectives of this pilot study were to: (1) assess data quality aspects at both the app and user levels, (2) assess the quantitative impact of ODK Scan on the data management processes, and (3) collect feedback on application user experiences to ascertain app usability, user interface design and workflow changes.

2. Material and Methods

2.1. Description of Technology

Open Data Kit (ODK) Scan is an Android-based mobile application and a part of the ODK 2.0 suite of tools [9,19]. It uses the device's built-in camera to capture an image of a scan-compatible paper form, which it feeds into a series of image processing algorithms (SURF, fast approximate nearest neighbor, and RANSAC) to establish a point-by-point mapping between the template and the captured image [20–23]. Using this mapping, Scan segments the form into small image snippets that each contains a single data field from the original form [24]. The image snippets fall into three possible prompt types: hand-filled bubbles, structured number boxes, and QR codes. Hand-filled bubbles are multiple-choice questions that are completed by filling in the bubble that corresponds with the desired response. Scan classifies hand-filled bubbles using a Support Vector Machine (SVM) and principal component analysis (PCA) [20,25,26]. Structured number boxes are a 2×3 grid of dots that can be connected by straight lines to draw numerals. Scan classifies numbers using a Multi-layer Perceptron (MLP) with back propagation of error [27]. Finally, QR codes are two-dimensional barcodes that can be processed as alphanumeric strings. Scan uses the Zebra Crossing (Zxing) library to classify QR Codes [28]. After Scan has classified each prompt, it inserts the data into an editable record in the ODK 2.0 database. Each digitized field includes a copy of its associated image snippet from the original form, which can be used to verify digitized fields. Once the record is in the ODK 2.0 database, users can use the other tools, such as ODK Survey and ODK Tables to view, edit, verify, aggregate, and sync the data or create customized services, interactions and reports with the data [19].

ODK Scan does not process freehand written text. Prompts that require written text will include boxes on the paper forms that users can fill with their hand written text, but Scan will leave the corresponding field in the database record blank. Hand written text fields can be manually transcribed on the device by viewing the prompt in ODK Survey, which will display the corresponding image snippet, and typing in the text via the phone's keyboard. Alternatively, they can be manually transcribed on a PC after the record has been synced, again by viewing the image snippet.

Scan and the rest of the ODK 2.0 tools do not require a consistent Internet connection to function [19]. All of the image processing is performed by the device itself, and the data is stored in a local database. An Internet connection is only required for syncing records with the cloud, and this can be done at the leisure of the user.

2.2. ODK Scan Workflow

Before ODK Scan is to be used, a scan-compatible form must be designed using the form designer. Paper copies of the form image are printed out and the form's definition files are copied to the Android-based device. After the form is filled, the user takes a picture using the built-in camera of

TheGlobalFund **Public-Private Mix (PPM) TB DOTS - TB 03** **MercyCorps**

Date of Registration:	Patient Name:	Mobile phone number for contact:							
Patient Identifier Number <input type="radio"/> A <input type="radio"/> B <input type="radio"/> C <input type="radio"/> D <input type="radio"/> E <input type="radio"/> F <input type="radio"/> G <input type="radio"/> H <input type="radio"/> I <input type="radio"/> J <input type="radio"/> K <input type="radio"/> L <input type="radio"/> M <input type="radio"/> N <input type="radio"/> O <input type="radio"/> P <input type="radio"/> Q <input type="radio"/> R <input type="radio"/> S (District code) (Health facility code) (Patient number) (Year)									
CNIC Number:	Sex: <input type="radio"/> Male <input type="radio"/> Female <input type="radio"/> TG	Name of TB Care Facility (PPM):							
Treatment Start Date (DD/MM/YY)	New Case (fill if Yes)	Transferred in (fill if Yes)	History of Previous Treatment (choose one option) <input type="radio"/> Relapse <input type="radio"/> Treatment after failure <input type="radio"/> Treatment after loss to follow-up <input type="radio"/> Other/Unknown history	Site <input type="radio"/> P <input type="radio"/> EP	Treatment Category <input type="radio"/> CAT-1 <input type="radio"/> CAT-2				
Patient Address:			Code of TB Care Facility (PPM):						
Smear (S), or Xpert MTB/RIF (X) results and other Examinations									
i. SMEAR									
'0' Month	Date of test/microscopy	Months 2/3	Date of test/microscopy	Month 5	Date of test/microscopy				
Result: (choose one option)	Negative 1-9 (Sputum) + (10-49 AFB) ++ (1-10 AFB/HPP) +++ (1-10 AFB/HPP)	Result: (choose one option)	Negative 1-9 (Sputum) + (10-49 AFB) ++ (1-10 AFB/HPP) +++ (1-10 AFB/HPP)	Result: (choose one option)	Negative 1-9 (Sputum) + (10-49 AFB) ++ (1-10 AFB/HPP) +++ (1-10 AFB/HPP)				
	<input type="radio"/> T <input type="radio"/> RR <input type="radio"/> TI <input type="radio"/> N <input type="radio"/> I		<input type="radio"/> T <input type="radio"/> RR <input type="radio"/> TI <input type="radio"/> N <input type="radio"/> I		<input type="radio"/> T <input type="radio"/> RR <input type="radio"/> TI <input type="radio"/> N <input type="radio"/> I				
ii. XPERT									
'0' Month	Date of test/microscopy	Months 2/3	Date of test/microscopy	Month 5	Date of test/microscopy				
Result: (choose one option)	T RR TI N I	Result: (choose one option)	T RR TI N I	Result: (choose one option)	T RR TI N I				
iii. CULTURE									
<table border="1"> <tr> <td>Treatment Outcome with Date</td> <td><input type="radio"/> Cured <input type="radio"/> Treatment Completed <input type="radio"/> Treatment Failure <input type="radio"/> Died <input type="radio"/> Loss to Follow Up <input type="radio"/> Not Evaluated</td> <td>Referred for Second Line <input type="radio"/> yes <input type="radio"/> no</td> <td>Remarks</td> </tr> </table>						Treatment Outcome with Date	<input type="radio"/> Cured <input type="radio"/> Treatment Completed <input type="radio"/> Treatment Failure <input type="radio"/> Died <input type="radio"/> Loss to Follow Up <input type="radio"/> Not Evaluated	Referred for Second Line <input type="radio"/> yes <input type="radio"/> no	Remarks
Treatment Outcome with Date	<input type="radio"/> Cured <input type="radio"/> Treatment Completed <input type="radio"/> Treatment Failure <input type="radio"/> Died <input type="radio"/> Loss to Follow Up <input type="radio"/> Not Evaluated	Referred for Second Line <input type="radio"/> yes <input type="radio"/> no	Remarks						

Figure 2. Scan-compatible TB03 form used in ODK Scan paper-to-digital system displaying fill-in bubbles, number boxes and text fields.

scan_TB03_Register1 back next

http://localhost:8635/tables/data/tables/scan_TB03_Register1/instance

Treatment start date

29/07/16

(a)

scan_TB03_Register1 back next

http://localhost:8635/tables/data/tables/scan_TB03_Register1/instance

Patient sex

male female TG

(b)

scan_TB03_Register1 back next

http://localhost:8635/tables/data/tables/scan_TB03_Register1/instance

Patient Name

Hamza Ali

Hamza Ali

(c)

scan_TB03_Register1 back next

http://localhost:8635/tables/data/tables/scan_TB03_Register1/instance

District

Hafizabad

(d)

Figure 3. Formats of data fields on ODK Scan-compatible TB03 form. (a) Machine-readable structured number boxes; (b) machine-readable fill-in bubbles; (c) user transcribed text field; (d) machine-readable QR code.

2.4. Design

A sequential, explanatory mixed-method research approach was employed to elucidate the technology use for an effective implementation of the ODK Scan paper-to-digital system. However, the mixed method research approach involves combining the quantitative analysis with qualitative reasoning. This research approach includes application of three methods: (1) quantitative assessment of the data quality at both app and user levels; (2) quantitative measurement of the time spent in filling forms, data transfer, aggregation and verification; and (3) qualitative feedback from the field workers (users) to capture their complete experience. However, each of these methods is explored in more detail below.

(1) Quantitative assessment of the data quality was done at both the app and user levels. At the application level, accuracy of ODK Scan's ability to recognize fill-in bubbles and number boxes was measured by comparing the raw Scan values (the digitized values which Scan's algorithm produced before human verification) with the data appearing in image snippets of the corresponding data fields. To calculate this, the team hired a Seattle USA-based data analyst to examine each individual number and fill-in bubble image snippet, to record the observed "ground truth" value as observed from the image, and to compare it with the corresponding raw value that ODK Scan processing recorded. Comparison of these field value pairs allowed us to calculate ODK Scan's accuracy rates for all the number and fill-in bubble fields.

At the user-level, the workflow allowed users to improve the quality of the data through verification and editing data fields manually. User-level data quality was assessed by comparing the same "ground truth" values for each field to the final values that the field worker synced with the server. These accuracy rates were predicted to be higher than the app-level accuracy rate because the field workers had the opportunity to correct errors from the "raw" Scan processing.

Furthermore, we calculated the accuracy rates of the existing (non-ODK Scan) manual entry processes by comparing the field values observed from the hardcopy TB03 register pages to the values manually entered by the data entry worker in the Microsoft Excel datasheets. This would provide us with a baseline for manual entry accuracy rate. Accuracy rates for both user-level digitization processes were calculated based on the full field value, rather than by individual digit or bubble (binary assessment of correct or incorrect for each field).

(2) Quantitative measurements of the time spent in filling forms, data transfer, aggregation and verification were recorded in the routine data management process using direct observation. They were then compared with the endline measures for the same data management processes using the ODK Scan system.

(3) Gathering of qualitative feedback on app use experience was done through surveys and interviews. The survey captured Likert-scale assessment, while face-to-face interviews using a semi-structured questionnaire were carried out with each of the field workers. Their experiences with app usability and workflow modification were explored to assess the effectiveness of the ODK Scan paper-to-digital system from a usability perspective.

3. Findings

3.1. Data Quality

App-level quality measure: During an intervention period of three months, 122 patients' paper-based records were scanned, digitized and checked for quality. All of these data were synced and available on the cloud-based ODK aggregate server for access and remote analysis. The first task was to assess ODK Scan's ability to correctly parse the user filled machine-readable data fields, the structured number boxes and the fill-in bubbles. Accuracy rates for these fields were calculated by manually reviewing the 122 records, consisting of 10,600 digits and 8000 fill-in bubbles, and comparing Scan's raw predicted value against the ground truth value of the image snippet. The recognition accuracy for the fill-in bubbles was found to be much higher than that of the number boxes. Including blank fields

(fields that users did not fill in), the Scan app classified 99.2% fill-in bubbles and 79.4% digits correctly (Table 1).

Table 1. Measurement of accuracy rates for recognition of fill-in bubbles and number boxes.

Type of Field	All Fields	Blank Fields Excluded
Number	79.4%	77.6%
Fill-in Bubbles	99.2%	97.2%

The misclassified number fields were further investigated by mapping accuracy measures of the number boxes on the TB form to specific fields on the physical paper. It was found that fields located on the right side of the page, the side farthest from the register book's binding, suffered from lower accuracy measures than the number boxes which were located closer to the center or left-hand side (Figure 4). Inspection of the captured images showed that the pages were bending away from the binding. This bending warped the point-to-point mapping of the captured image to the template file, which caused the expected image snippet coordinates to mismatch. The resulting image snippets were misaligned, meaning that there would be missing edges and other important information. The structured number classifier's accuracy suffered when attempting to classify misaligned image snippets.

The figure displays a composite of a Public-Private Mix (PPM) TB DOTS - TB 03 form. The form is divided into several sections:

- Top Section:** Date of Registration, Patient Name, Mobile phone number for contact, CNIC Number, Sex, Age, Name of TB Care Facility (PPM), District.
- Patient Identifier Number:** A grid of 12 boxes labeled A through S, with a breakdown: A (1), B (1), C (1), D (1), E (1), F (1), G (1), H (1), I (1), J (1), K (1), L (1), M (1), N (1), O (1), P (1), Q (1), R (1), S (1). Accuracy: 84.5%.
- Treatment Start Date:** A grid of 12 boxes. Accuracy: 84.6%.
- History of Previous Treatment:** Options include Relapse, Treatment after failure, Treatment after loss to follow-up, Other/Unknown history. Accuracy: 84.0%.
- Site and Treatment Category:** Site (O, P, EP), Treatment Category (CAT-1, CAT-2).
- Smear (S), or Xpert MTB/RIF (X) results and other Examinations:**
 - i. SMEAR:** Grids for '0' Month, Months 2/3, Month 5, and End of Treatment. Results are categorized by sputum sample count (1-9, 10-99, >100) and microscopy type (AFB, AFB+HPF, AFB&HPF). Accuracy: 82.8%, 95.0%, 87.9%, 59.1% respectively.
 - ii. XPERT:** Grids for '0' Month, Months 2/3, Month 5, and End of Treatment. Results are categorized by sample type (T, RR, TI, N, I). Accuracy: 87.6%, 93.1%, 83.1%, 51.4% respectively.
 - iii. CULTURE:** Grids for Treatment Outcome with Date (Cured, Treatment Completed, Treatment Failure, Died, Lost to Follow Up, Not Evaluated) and Referred for Second Line (yes/no). Remarks section shows Accuracy of Digits in Field: 90-100% (green), 80-89% (yellow), 50-79% (red).

Figure 4. Mapping of the accuracy measures of the number boxes.

User-level quality measure: The ODK Scan workflow includes viewing, editing and verifying data before it is synced with the server. Field workers were therefore given complete responsibility for the quality assessment and improvement. However, in the existing manual data entry process, this responsibility is shared between a field worker and a data management officer (DMO). The DMO is responsible for data entry and quality at the next level of management. When data quality (accuracy) of synced records were checked and compared with the quality of manually entered records, it was found that the existing manual data entry process resulted in a data accuracy rate of 94%, while the ODK Scan process only reached 86%. As described in the methodology section, both of these accuracy

rates were calculated by comparing values observed from the paper records with the values that were saved to the digital record in each process. While it is understandable that human accuracy for data entry did not reach a perfect 100% for either processes, the discrepancy between the two rates shows that the task-dependence of field-worker for data verification with ODK Scan did not meet the quality of the existing manual data entry. This discrepancy between the manual data entry and verified ODK Scan data accuracy rates is further explored in the discussion section.

3.2. Impacting Data Management Processes

In the TB control program, the routine data management process begins with the collection of data from all enrolled general practitioners (GPs). Once the consolidated paper-based report is prepared it is transferred to the next level of management (sub-recipient office), which then passes through multiple desk reviews and is finally entered into an Excel spreadsheet. The digitization of paper-based records takes a significant amount of time before it is shared with the next level of management (principle recipient office) for finalization and progress review meetings (Figure 5). Time spent in form-filling, data collection, transfer, aggregation and verification was measured by the direct observation. At baseline, these measures were recorded by observing the routine data management process. However, at the endline, the same measures were recorded by observing the ODK Scan workflow.

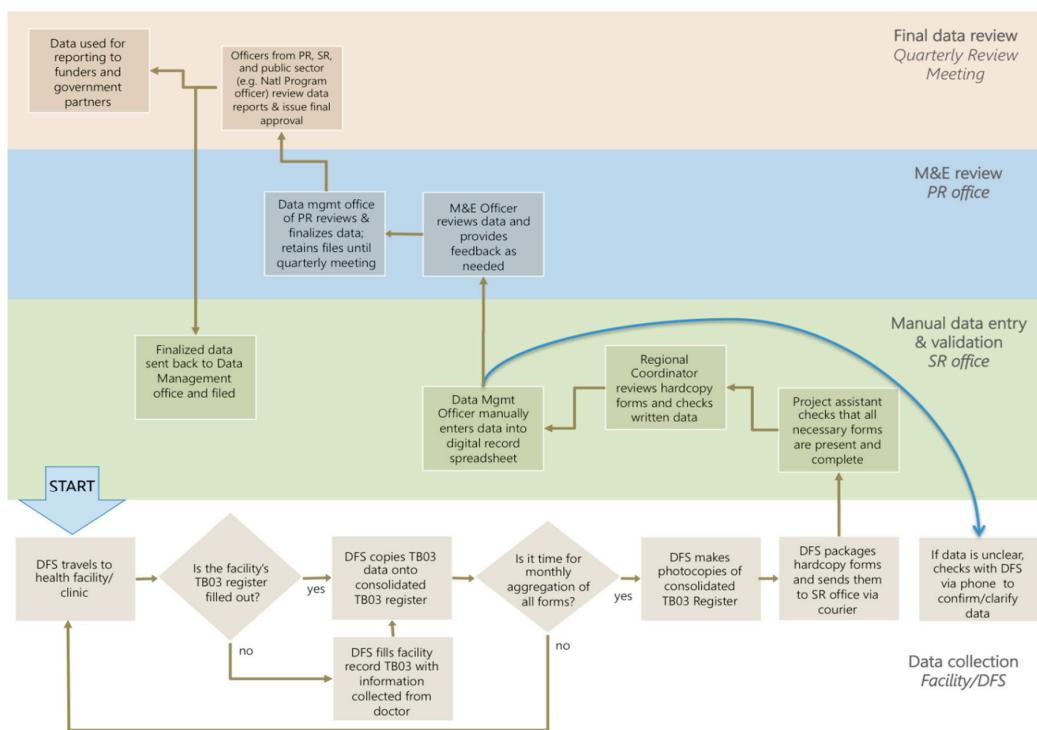


Figure 5. Routine data management process in the TB (Tuberculosis) control program of Pakistan.

Filling in the ODK scan-compatible TB form took longer than the old form. Extra attention was required for filling in bubbles and writing structured numbers on the guiding dots correctly. Each field worker ($n = 4$) was observed while filling in three Scan-compatible forms. The average time taken to fill 12 Scan-compatible forms was 02 min 57 s (max, 03:04 s; min, 01:48 s). However, on average it took 1 min and 27 s to fill in one patient record using the existing TB03 register. The next aspect of data management, data collection, was not affected by the ODK Scan workflow because the process remained the same: visit all of the enrolled GPs of a district. Finally, with the routine data management process, once data collection was completed, it took 1–2 days to transfer data via local courier service, in the form of hard-copies. However, with the ODK Scan process the same transfer of data only lasted

a few min. It only required the user to get to an area of internet connectivity and send the data to the ODK Aggregate Server.

In the routine data management process, aggregation of records required manual data entry into a spreadsheet, and digitization of one record took 2 min. This digitization process was reduced to 10 s in the ODK Scan workflow. Moreover, the data verification process was greatly altered, as in the ODK Scan process the field worker took the sole responsibility of viewing and verifying data, while in the routine data management process this responsibility was shared between field worker and DMO. In terms of time, verification of paper based records took 20 s per record for the DMO, while in the ODK Scan process it took 3 min per record for the field worker. The increased consumption of time in the verification step of the ODK Scan is because of the app workflow involving identification and rectification of the misrecognized machine-readable data fields by moving point-by-point. The ODK Scan process provided a significant time reduction for the data transfer and aggregation steps; however, time required for the verification and form filling were increased as compared to corresponding processes of the routine data management (Table 2, Figure 6).

Table 2. Impact of ODK Scan on various processes of Data Management.

Data Management Processes	Routine Data Management Process	ODK Scan Process	Time Consumed in Redesigned Process
Form Filling (per form)	1 min 27 s	2 min 57 s	↑
Data Collection (complete)	3–4 days	3–4 days	↔
Transfer (complete)	1–2 days	3 min	↓
Aggregation (per record)	2 min	10 s	↓
Verification (per record)	20 s	3 min	↑

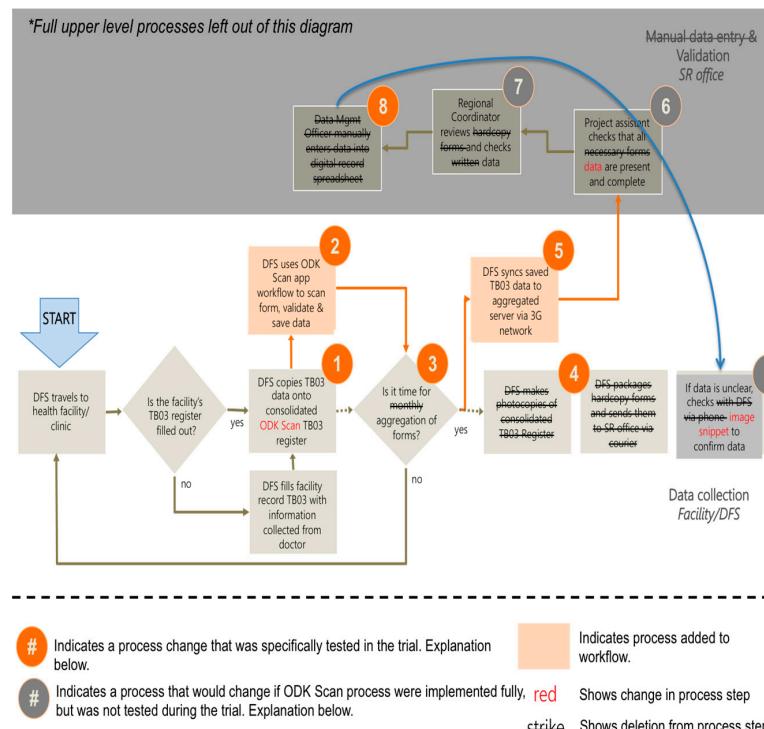


Figure 6. Redesigned data management process in ODK Scan paper-to-digital system.

The redesigned data management process saved resources by eliminating tasks that demanded time and money. In the routine data management process, paper-based records were sent to the office for digitization via local courier service. However in the redesigned process, manual entry of the

records at the office is completely eliminated and data is sent to the server electronically through internet connectivity. The redesigned data management process offered flexibility to supervisors in reviewing digital data and enabled them to provide feedback immediately.

3.3. Feedback on App Usability and Workflow Modifications

The field workers learned quickly how to use the smartphones and swiftly became comfortable with the application's workflow: most of them were already using smartphones in their personal lives. Nevertheless, the implementation of ODK Scan was guided by the principles of minimal disturbance in current workflow and generating value for users to improve the chances of adoption. Face-to-face interviews were conducted independently with each of the field workers at the end of the intervention period. These interviews were supported by a semi-structured questionnaire, which included hints to guide the interviewer to ensure the completeness of the feedback. The interviews were conducted by a research team member in the language of the enrolled field workers (Urdu). All interviews were digitally recorded, transcribed in Urdu and then translated into English. The interview recordings and the corresponding translations were then checked by another team member for consistency and completeness. Data is segregated based on the various steps of the app workflow for the ease of analysis and presentation.

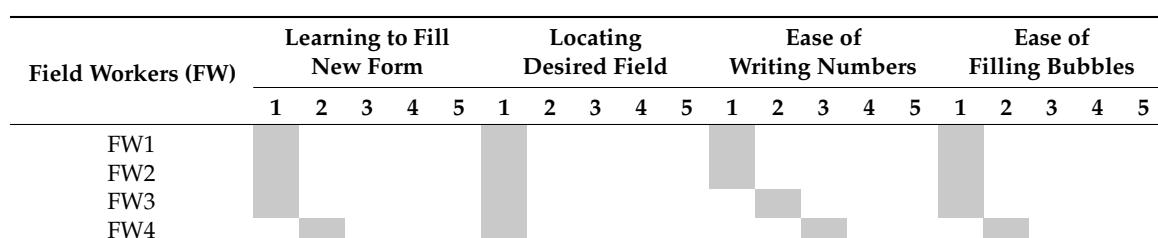
3.3.1. Form Filling

In both routine and redesigned data management processes users are required to fill paper-based form. The form designs used in routine and redesigned processes are completely different, but data fields are common between the two designs. Furthermore, the Scan-compatible TB03 form allowed more structured and fewer hand written text fields, whereas the standard TB03 form included hand written data fields only. The ODK Scan process begins with filling in the ODK Scan-compatible form. This includes marking the correct bubble for the fill-in bubble fields, linking guide dots for the structured number fields and writing in the text fields. Field test results showed that this step took more time as compared to the corresponding step of the routine data management process. Moreover, in terms of learning to fill in the new form, all of the field workers adapted to the form well and found it easy to locate the desired field. Nonetheless, almost all field workers were unhappy about the time-consuming aspect of the form filling for the ODK Scan process (Table 3), while at the same time they acknowledged the benefit of using number boxes and fill-in bubbles. One DFS said, "This experience was very good. This is excellent. Especially dots and bubbles have eliminated any chance of the data management officer contacting us for clarification or to address any readability issue."

Interestingly, one DFS highlighted a different aspect of this, saying

"[I do not like] filling the dots on the form as they are time consuming. Also, filling the dots embarrasses me whenever I do that in the presence of others. It seems like I am a kid and practicing handwriting."

Table 3. Self-administered questionnaire items.



1 corresponds to very easy and 5 corresponds to difficult.

3.3.2. Scanning Form

Scanning a Scan-compatible form requires the user to take a picture of the paper form using smartphone's built-in camera. This lets the application process the form in the background and detects machine-readable, hand-marked data fields. Once the processing of image is finished, a notification appears in the top toolbar. Field workers understood the importance of taking a good quality picture, and one described it as follows:

"A good picture is one which is not blurred, and properly focuses the form and covers all the edges of forms. No corner should be cut."

Field workers considered this step important and easy to learn. Knowing this step to be important for having accurate and better classification results, they preferred to take image at a relatively calmer place. One field worker said, "Because this is delicate task and to make its editing error-free, I conduct this exercise at home."

3.3.3. Verifying Data

When an image is processed, the record is stored into the ODK 2.0 database. The user then launches the ODK Survey app by pressing the "Transcribe" button to verify the data. This allows the field worker to move field-by-field checking for any necessary edits. Most of the field workers considered the data verification process time-consuming, which may have contributed to the relatively lower data quality in comparison to the manual data management process. One field worker explained this process as follows: "I performed verification by pressing next option to move from one field to another. I checked all data points of one form and verify each field to confirm that data written on form is properly scanned and all mistakes are rectified. The information is easily viewable."

Moreover, one field worker explained the positive aspect of this verification step and said, "Using this app feature, I feel more confident about the quality of data as I view and verify each data fields one-by-one. It feels like I am taking complete charge of the data quality."

3.3.4. Sending Data

Sending data to the next management level was significantly easier with the ODK Scan process. In the Scan process, this simply involves getting to an area with internet connectivity and sending data to the server by syncing it. Conversely, in the routine data management process, it is very time and energy consuming: copies of the TB register must be sent to the office through a reliable courier service. One of the field workers explained this process: "Using sync feature is easier because I do editing mostly at home then I reach an area of connectivity and sync forms. While, for sending paper forms via courier, first I have to carry big sized TB03 register to photocopier to make the copies of TB03, then I have to visit courier office to send the parcel. It will not reach the same day, at least it takes one day and then I have to call office to confirm if they have received the parcel/documents."

Another field worker explained the process of data sharing that he adopted over the intervention period: "Once editing is finalized, then we connect device to Internet—mostly to WiFi—and open Core Services app for syncing. First I verify my email address then syncing is confirmed. Mostly I used to sync the data from public partner's office, and if I am far away then I go to local Internet café to sync data. To ensure my connectivity with Internet I check and confirm WiFi icon on top of the screen."

However, dependence on WiFi internet connection for syncing the data was challenging for two of the field workers, and one of them said, "It would be very great and easy for us if ODK can sync forms via 3G Internet connectivity. As it is, I scan and edit the forms at home and for syncing forms I have to travel to place where I can have access to WiFi which takes too much time."

3.3.5. Report Generation

Regular sputum smear microscopy is an important medical aspect in ensuring adherence to the treatment protocol and declaring the treatment outcome of a TB patient. It is extremely difficult to

review paper-based records and prepare a list of patients whose sputum smear microscopy is due in the next month. This time-consuming exercise is iterative and considered as a fundamental responsibility of the field workers for tracking treatment progress of the registered patients. With the ODK Scan process, generating this actionable information became easier by running a customized follow up visit report on the smartphone in offline mode (Figure 7). This enabled field workers to keep a constant check on the treatment progress of the patients registered in his area of control. Moreover, all field workers talked very positively about the benefits of this report. One of them said, "My existing practice is to review TB03 in the beginning of month to identify patients with follow up appointments. This task takes lot of time and mental energies as it is difficult to develop such list. With that view point, if I see this application, this is an excellent idea that I open app and all follow up visits will be appeared."

Others said:

"This is the best feature of this application, as now we know in advance when we have to call the patient for follow-up."

"Previously my practice was to visit all the GPs to enlist patients who have follow up visit due or overdue. Now with this report I do not have to visit the GPs to know patient's follow up visit date."

"I would say that, follow up visit report has revolutionized our work. Now, we will never miss out any patient and upon click we view all needed information required for follow up."



Figure 7. Process of Follow up Visit Report (left to right).

3.3.6. Suggestions for Improvement

After analyzing the content of interviews, it was evident that the field workers completely understood the app workflow and appreciated the benefits of using ODK Scan in improving their work efficiency and performance. Moreover, field workers were eager to continue application use. In addition to simple aggregation reports, field workers wanted to enhance the features of the application. For instance, they suggested including a bookmark feature to tag any patient record, if he/she is not accessible via telephone at the first attempt. With the number of the records available in a table on the user's device, the most popular requirement was the patient look-up feature. As TB treatment involves multiple visits of the same patient for medical checkup, it is becoming increasingly difficult for field workers to retrieve a relevant patient's record from the ODK Tables default interface upon patient's revisit. A look-up feature will help field workers to avoid re-scanning and re-verifying forms, and the follow up service will be updated in an existing patient's record.

Currently, synchronization involves the movement of data in either direction. This means new records on the device are sent to the server and at the same time new records available at the server, synced by others, are downloaded onto the user's device. Deleting a record, either deliberately or mistakenly, is a potential threat of data loss. Therefore, defining user permission attributes to restrict data access to only the appropriate individuals would improve the Scan workflow. Lastly, one of the field workers suggested that instead of scanning, a direct-to-digital solution seems more feasible.

4. Discussion

This paper compares paper-based and technology-assisted data management processes and tries to extrapolate the value of mHealth solutions for capturing remote health data. In particular, the present study assesses the usability and technological aspects of the ODK Scan supported by the perspective of the field workers in relation to their work settings. However, it is also noted that in the developing country context, a prerequisite to the success of mHealth interventions is the consideration of developers and implementers about the field workers' expectations, needs and work settings [29,30]. Moreover, a gap between the work practices and use of the new technologies is attributed by the poor implementation design of the technology-based interventions [31].

In the Public-Private Mix model of the TB control program of Pakistan, despite available direct-to-digital data collection solutions, the ODK Scan paper-to-digital system was chosen because of the current healthcare system requirement. Largely, the healthcare system relies on paper-based records for recording, reporting, verification, validation and other purposes. This setting informed the selection of ODK Scan paper-to-digital system. In other mHealth interventions where Android-based smartphones are used primarily for data collection, the ability of smartphones to transfer data in real time from remote areas to centrally located offices is achieved by exploiting the internet connectivity [15,32]. Similarly, in this field test, internet connectivity benefited field workers in transferring the data and labor involved in sharing the data in the form of hard copies was avoided. However, the ability of smartphones to transfer data by using 3G internet connectivity instead of WiFi would be highly beneficial. More than just digitization of patient's records, technology also supports report generation, case management and patient follow-up [33].

Data quality achieved through the ODK Scan process is not impressive if compared with its counter process. This drop in quality can be explained through imperfect computer vision algorithms, and the change in the field workers roles; previously field workers were not solely responsible for ensuring data quality. In terms of time, the routine process of data review takes much longer than the new process of the data review and verification. Therefore, an alternative approach to improve the data quality would be to carry out supportive supervision, so that acceptable level of quality of the final set of data could be achieved [34]. However, supportive supervision role can be strengthened by providing appropriate training and timely feedback to the field workers on the data quality aspects. Additionally, effective data verification approaches require determination of the data quality characteristics and the application of quality standards [35]. However, the benefits of the mHealth technologies are improved with increased validity of the digital data [36]. Data quality could also be enhanced by improving ODK Scan's algorithms, but this is an expensive and difficult proposition. The worst of the accuracy problems were associated with poor form alignment due to bending pages. This might be helped by improving the form alignment algorithm or training the classifiers to handle misaligned snippets. Classification accuracy might also be improved by using context information, such as specified valid numeric or date ranges, though this would increase form design complexity.

In the ODK Scan pilot test, unrestricted mobile and internet usage was authorized in good faith to generate ownership and responsibility among the field workers. It is also reported that the free use policy has generated a sense of ownership and empowerment among the health workers by recognizing the significance and convenience of the smartphone usage in their practices [37].

The most challenging aspect of the use of ODK Scan was filling in Scan-compatible forms that included fill-in bubbles, number boxes and text fields. Additionally, this field test recorded significant

time consumption on data verification and editing processes, which is expected to be reduced with field workers becoming more accustomed with scan-compatible form and related processes. Likewise, one of the data field workers also mentioned the practice aspect of the time reduction during an interview. On one hand, viewing and verifying all processed data values is beneficial for ensuring data quality, but on the other hand this requires the field worker to spend more time finalizing the scanned record after correcting Scan's errors. Therefore, more errors in the machine-readable data fields resulted in the frustration of field workers. Furthermore, lower recognition accuracy of the number boxes indicates a need for additional application engineering work to make ODK Scan more practical and workable solution. However, it is also an essential consideration for an implementing organization to include as many machine-readable data fields as required [38]. Additionally, the opinion of the field workers must be included during the form design process, so that the workload is anticipated and subsequently accepted by them and only important data fields are chosen for the ODK scan paper-to-digital system. Complexity of the mHealth interventions requires change management considerations, including changes in the behavior of healthcare providers and professionals, and systems and processes in delivering care [39,40].

Considering the short duration of the ODK Scan pilot phase, its impact on the TB case management aspects and treatment outcomes could not be measured. After feedback-informed technology enhancements, its expansion plan in the TB control program of Pakistan will be prepared. In current mHealth domain, a key challenge is to make strong evidence available for advocating the proliferation of mHealth use in improving health outcomes. Therefore, in future studies, we aim to measure ODK Scan's effectiveness in ensuring fidelity to treatment protocols and improving treatment outcomes.

5. Conclusions

Although it is too early to demonstrate a direct link between digitization of data collection and improvement in health outcomes, mHealth technologies enable public health professionals and program managers to make data available for use in an easy and timely fashion. Prompt generation of actionable information from the digital records is major advantage of the technology. However, it is important to integrate application use within the routine workflow, and providing minimal disturbance of existing practices and continuous feedback are important factors for technology acceptance among the field workers. Efforts made to enhance technology features and improving users' experiences are important considerations for the successful implementation of mHealth interventions in the developing countries. However, a gradual adoption is necessitated by human resource factors and the pace of technological advancements. Hence, it is equally important to devise a national-level policy, or at least organizational policy, for the integration of the mHealth technologies into public health programs to maximize opportunities to reach needy communities.

Acknowledgments: This paper was put together by the Monitoring, Evaluation and Learning Unit team of the Mercy Corps (MC) Pakistan. On behalf of the MC Pakistan, the authors would like to offer their thanks for the excellent support given by the technical teams of Village Reach and the University of Washington for collaborative mHealth work in the Public-Private Mix Model of the TB control program of Pakistan. The authors would also like to thank Aamna Rashid for her support and the participating field workers, Mudassar, Naveed, Shuja and Zuhair, for their patience, support and feedback.

Author Contributions: S.M.A. prepared the final draft of the manuscript and contributed to the development of the analysis plan. R.P. developed the overall design of the study, conducted analysis of the quantitative data and developed graphical presentations. She also developed the data collection tool. A.N. assigned resources for this field test and provided managerial oversight. F.N. and N.A. conceived the idea and prepared the analysis plan, implementation design and analyzed qualitative data. M.I. conducted interviews, translated them into English and managed field activities. M.I. and J.A. conducted the literature search. J.A. managed the technical aspects of the application and also helped M.I. in translating interviews. Technical and development work was led by J.B., R.A. oversaw the development of technology and checked the final manuscript for consistency and scientific rigor. All authors read and approved the final manuscript.

Conflicts of Interest: R.A. and J.B. are involved in the development and management of ODK, an open source platform for data collection.

References

1. World Health Organization (WHO). *Global Tuberculosis Report-2015*; WHO: Geneva, Switzerland, 2015; Available online: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf (accessed on 27 June 2016).
2. Sibley, L.; Dennis, M.; Sarfas, C.; White, A.; Clark, S.; Gleeson, F.; McIntyre, A.; Rayner, E.; Pearson, G.; Williams, A.; et al. Route of delivery to the airway influences the distribution of pulmonary disease but not the outcome of *Mycobacterium tuberculosis* infection in rhesus macaque. *Tuberc. J.* **2016**, *96*, 141–149. [CrossRef] [PubMed]
3. World Health Organization (WHO). Available online: <http://www.emro.who.int/pak/programmes/stop-tuberculosis.html> (accessed on 27 June 2016).
4. World Health Organization (WHO). *Understanding and Using Tuberculosis Data*; WHO: Geneva, Switzerland, 2014; Available online: http://apps.who.int/iris/bitstream/10665/129942/1/9789241548786_eng.pdf?ua=1&ua=1 (accessed on 27 June 2016).
5. World Health Organization (WHO). *A Guide to Monitoring and Evaluation for Collaborative TB/HIV Activities*; WHO: Geneva, Switzerland, 2015; Available online: http://apps.who.int/iris/bitstream/10665/150627/1/9789241508278_eng.pdf?ua=1 (accessed on 27 June 2016).
6. World Health Organization (WHO). *Digital Health for End TB Strategy: An Agenda for Action*; WHO: Geneva, Switzerland, 2015; Available online: http://www.who.int/tb/areas-of-work/digital-health/Digital_health_EndTBstrategy.pdf (accessed on 27 June 2016).
7. Division of Tuberculosis Elimination. *Treatment of Drug-Susceptible Tuberculosis Disease in Persons Not Infected with HIV*; Centers for Disease Control and Prevention (CDC): Atlanta, GA, USA, 2012. Available online: <http://www.cdc.gov/tb/publications/factsheets/treatment/treatmenthivnegative.htm> (accessed on 27 June 2016).
8. Timimi, H.; Falzon, D.; Glaziou, P.; Sismanidis, C.; Floyd, K. WHO guidance on electronic systems to manage data for tuberculosis care and control. *J. Am. Med. Inform. Assoc. (JAMIA)* **2012**, *19*, 939–941. [CrossRef] [PubMed]
9. Dell, N.; Perrier, T.; Kumar, N.; Lee, M.; Powers, R.; Borriello, G. Paper-digital workflows in global development organizations. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing, Vancouver, BC, Canada, 14–18 March 2015; pp. 1659–1669.
10. Agarwal, S.; Rosenblum, L.; Goldschmidt, T.; Carras, M.; Goel, N.; Labrique, A.B. Mobile Technology in Support of Frontline Health Workers: A Comprehensive Overview of the Landscape, Knowledge Gaps and Future Directions. John Hopkins University Global mHealth Initiative 2016. Available online: https://dl.dropboxusercontent.com/u/5243748/mFHW%20Landscape_2016%20Final.pdf (accessed on 27 June 2016).
11. Sa, J.H.G.; Rebelo, M.S.; Brentani, A.; Grisi, S.J.F.E.; Iwaya, L.H.; Simplicio, M.A.; Carvalho, T.C.M.B.; Gutierrez, M.A. Georeferenced and secure mobile health system for large scale data collection in primary care. *Int. J. Med. Inform.* **2016**, *94*, 91–99. [CrossRef] [PubMed]
12. Hamine, S.; Gerth-Guyette, E.; Faulx, D.; Green, B.B.; Ginsburg, A.S. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review. *J. Med. Internet Res.* **2015**, *17*, e52. [CrossRef] [PubMed]
13. Royal College of Nursing. *Tuberculosis Case Management and Cohort Review: Guidance for Health Professionals*; RCN: London, UK, 2012; Available online: https://www2.rcn.org.uk/__data/assets/pdf_file/0010/439129/004204.pdf (accessed on 5 October 2016).
14. Iribarren, S.J.; Rubinstein, F.; Discacciati, V.; Pearce, P.F. Listening to those at the frontline: Patient and healthcare personnel perspectives on tuberculosis treatment barriers and facilitators in high TB burden regions of Argentina. *Tuberc. Res. Treat.* **2014**. [CrossRef] [PubMed]
15. Denkinger, C.M.; Grenier, J.; Stratis, A.K.; Akkihal, A.; Pant-Pai, N.; Pai, M. Mobile health to improve tuberculosis care and control: A call worth making. *Int. J. Tuberc. Lung Dis.* **2013**, *17*, 719–727. [CrossRef] [PubMed]
16. Van Velthoven, M.H.; Car, J.; Zhang, Y.; Marušić, A. mHealth series: New ideas for mHealth data collection implementation in low- and middle-income countries. Editorial. *J. Glob. Health* **2013**, *3*, 1–3. [CrossRef] [PubMed]

17. Kaplan, W.A. Can the ubiquitous power of smartphones be used to improve health outcomes in developing countries? *Glob. Health* **2006**, *2*, 9. [[CrossRef](#)] [[PubMed](#)]
18. Mosa, A.S.; Yoo, I.; Sheets, L. A systematic review of healthcare applications for smartphones. *BMC Med. Inform. Decis. Mak.* **2012**, *12*, 67. [[CrossRef](#)] [[PubMed](#)]
19. Brunette, W.; Sundt, M.; Dell, N.; Chaudhri, R.; Breit, N.; Borriello, G. Open data kit 2.0: Expanding and refining information services for developing regions. In Proceedings of the 14th Workshop on Mobile Computing Systems and Applications, Jekyll Island, GA, USA, 26–27 February 2013.
20. Dell, N.; Breit, N.; Chaluco, T.; Crawford, J.; Borriello, G. Digitizing paper form with mobile imaging technologies. In Proceedings of the 2nd ACM Symposium on Computing for Development, Atlanta, GA, USA, 10–11 March 2012.
21. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]
22. Muja, M.; Lowe, D. Fastest approximate nearest neighbors with automatic algorithm configuration. In Proceedings of the International Conference on Computer Vision Theory and Application (VISSAPP'09), Lisboa, Portugal, 5–8 February 2009.
23. Fischler, M.; Bolles, R. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
24. Dell, N.; Breit, N.; Wobbrock, J.O.; Borriello, G. Improving form-based data entry with image snippets. In Proceedings of the 13 Proceedings of Graphic Interface, Regina, SK, Canada, 29–31 May 2013.
25. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
26. Pearson, K. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philos. Mag.* **1901**, *2*, 559–572. [[CrossRef](#)]
27. Robert, H.-N. Theory of the backpropagation neural network. *Proc. Int. Jt. Conf. Neural Netw.* **1989**, *2*, 65–93.
28. Zebra Xing (Zebra Crossing). Available online: <https://github.com/zxing/zxing> (accessed on 4 October 2016).
29. Paul, R.J.; Ezz, I.; Kuljis, J. Health information systems: A patient-user perspective. *Health Syst.* **2012**, *1*, 85–95. [[CrossRef](#)]
30. Shachak, A.; Montgomery, C.; Dow, R.; Barnsley, J.; Tu, K.; Jadad, A.R.; Lemieux-Charles, L. End-user support for primary care electronic medical records: A qualitative case study of users' needs, expectation, and realities. *Health Syst.* **2013**, *2*, 198–212. [[CrossRef](#)] [[PubMed](#)]
31. Ali, S.M.; Giordano, R.; Lakhani, S.; Walker, D.M. A review of randomized controlled trials of medical record powered clinical decision support system to improve quality of diabetes care. *Int. J. Med. Inform.* **2016**, *87*, 91–100. [[CrossRef](#)] [[PubMed](#)]
32. Earth Institute. *Barriers and Gaps Affecting Mhealth in Low and Middle Income Countries: A Policy White Paper*; mHealth Alliance: Washington, DC, USA, 2010; Available online: <http://cgasd.columbia.edu/files/2012/11/mHealthBarriersWhitePaperFINAL.pdf> (accessed on 15 July 2016).
33. Dwolatzky, B.; Trengove, E.; Struthers, H.; Mc-Intyre, J.; Martinson, N. Linking the global positioning system (GPS) to a personal digital assistant (PDA) to support tuberculosis control in South Africa: A pilot study. *Int. J. Health Geogr.* **2006**, *5*, 34. [[CrossRef](#)] [[PubMed](#)]
34. Luk, R.; Bancroft, E. Lesson Learned: ICTs for Supply Chain Management in Low-Resource Settings. Village Reach and Dimagi. 2013. Available online: <https://www.k4health.org/toolkits/mhealth-planning-guide/lessons-learned-icts-supply-chain-management-low-resource-settings> (accessed on 15 July 2016).
35. Divorski, S.; Scheirer, M.A. Improving data quality for performance measures: Results from a GAO study of verification and validation. *Eval. Progr. Plan.* **2001**, *24*, 83–94. [[CrossRef](#)]
36. Poku, A.M.; Katz, A.T. Development of mobile health capabilities for remote data collection in resource-limited settings. *Johns Hopkins APL Tech. Dig.* **2014**, *32*, 679–684.
37. Little, A.; Medhanyie, A.; Yebyo, H.; Spigt, M.; Dinant, G.; Blanco, R. Meeting community health worker needs for maternal health care service delivery using appropriate mobile technologies in Ethiopia. *PLoS ONE* **2013**, *8*, e77563. [[CrossRef](#)] [[PubMed](#)]
38. Dell, N.; Crawford, J.; Breit, N.; Chaluco, T.; Coelho, A.; McCord, J.; Borriello, G. Integrating ODK Scan into the community health worker supply chain in Mozambique. In Proceedings of the Sixth International Conference on Information and Communication Technologies and Development, Cape Town, South Africa, 7–10 December 2013; Volume 1, pp. 228–237.

39. The PLOS Medicine Editors. A reality checkpoint for mobile health: Three challenges to overcome. *PLoS Med.* **2013**, *10*, e1001395.
40. Tomlinson, M.; Rotheram-Borus, M.J.; Swartz, L.; Tsai, A.C. Scaling up mHealth: Where is the evidence? *PLoS Med.* **2013**, *10*, e1001382. [CrossRef] [PubMed]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).