Data accuracy:
measuring digitisation within Shifo solutions - has scanning technology overcome its troubling past?
This report is written by

Timothy Anderson ¹, Alice Titialii ², Malaika Mikaelsson ², Shahnoza Eshonkhojaeva ²

Authors’ affiliation: 1 - Tallinn University , 2 - Shifo Foundation

Cover photo: Shifo

This document can be freely used for educational and non-commercial purposes.
# Table of contents

Abstract......................................................................................................................................................4

1. Background ..............................................................................................................................................5

2. Methods..................................................................................................................................................5

3. Results....................................................................................................................................................7

4. Conclusion.............................................................................................................................................9

5. References ...........................................................................................................................................10
Abstract

**Background:** Health information systems are a vital part of healthcare delivery, and it is essential that the data presented is reliable. The aim of this study was to measure accuracy of the digitisation process in Shifo’s health information solutions, as part of a series assessing data quality.

**Methods:** Hand-recorded data from clinics was compared to digitised data found in the Smart Paper Technology database. Information was collected from three countries: Afghanistan, Uganda, and The Gambia.

**Results:** Findings from this study show that the digitisation of Smart Paper Forms has an accuracy rate of above 99% (Afghanistan: 99.40-99.60%, Uganda: 99.17-99.20%, The Gambia: 99.80%).

**Conclusion:** Results indicate that Smart Paper Technology Engine is highly accurate in its rendering of healthcare records. We expect data accuracy to further improve with incoming refinements to the software.
1. Background

Health information systems play a pivotal role in the strengthening of healthcare delivery (Teklegiorgis et al. 2016). A good health information system should provide decision-makers and health planners at all levels with reliable, high-quality data. Assessments of data quality are among the most important evaluations for health information systems, as the data produced needs to be correct and consistent if planners and decision-makers are expected to use it. Health facility data is vital information to have when assessing national health programmes and policies (WHO 2012); and inaccurate or low quality data within health care could negatively affect both costs for the health system and the quality of care provided to patients (Peabody et al. 2004).

Shifo’s health information solution hinges on the use of Smart Paper Forms to collect data at clinics and then digitising and processing the information using a software solution titled Smart Paper Technology Engine. The solution then creates a database of all health information where it can be easily searched and represented in different formats. Shifo’s solutions represent an alternative form of eHealth solutions, where the digitisation does not occur at point-of-care, but instead Smart Paper Forms are used to collect data during service delivery and digitisation occurs at a later stage when forms are scanned. Although there are significant advantages of this digitisation process, there has previously been concerns that deficiencies in scanning technology negatively impact data accuracy (see, for example, Kenny et al. 2017; Liu et al. 2017; Landis-Lewis et al. 2015). Can we ensure that the data in Smart Paper Forms are being read and processed correctly after scanning?

This study is one part in a series of studies assessing the quality of data produced using Shifo’s Smart Paper Technology solutions. When discussing the data quality of health information systems, the term ‘quality’ encompasses three parts: completeness, timeliness, and accuracy. This particular study will focus on the accuracy of data digitisation using Smart Paper Forms. This assessment is being undertaken to measure how much trust can be placed in the Smart Paper data produced, and to investigate potential areas in which the system can be further improved.

We seek to answer a key question regarding data accuracy within Shifo solutions:

- Is the data from patient documents (Smart Paper Forms) being accurately read and processed by the Smart Paper Technology Engine?

2. Methods

Data processing and verification within Shifo solutions

The Smart Paper Technology Engine (SPTE) comprises a suite of tools which facilitate the collection, scanning, recognition, automatic processing and generation of reports based on clinical and administrative data from health service delivery points. SPTE receives scanned files and automatically converts images of handwritten text and check marks/ticks from Smart Paper Forms into machine-encoded text. When the system detects unrecognisable data, such as bad handwriting, poor marking, and/or damaged/dirty document images, SPTE automatically sends it for review to a Data Verification Officer, who is equipped with a data verification station, to enable the control of data accuracy and verification. By viewing unrecognised images of handwritten text and/or hand-marked data, the Data Verification Officer rectifies and/or inputs actual data seen in the document image (Step 4 in Box 1). After the data has been reviewed, it is passed on for quality assurance. At the data quality assurance stage (Step 5 in Box 1), a Quality Assurance Officer performs audits and does further checks on the
automatically recognised information. He/she reviews document batches and the handwritten and marked data recognised automatically by SPTE. During the time of data collection for this study, The Gambia was the only country to have a new function introduced at the verification stage, which flags errors such as not ticking all boxes for vaccines that should be administered as a group. This new feature helps the Data Verification Officer identify and rectify unrecognised ticks/check marks.

This study used the data passed on for quality assurance from the Data Verification Officer, and compared the information with data found in the Smart Paper Forms.

**Data collection steps**

1. Data entry
2. Scanning
3. Processing
4. Verification
5. Quality assurance
6. Export

Box 1. An illustration of each data collection step within the Smart Paper Technology Engine

**Measuring digitisation**

This study measured one key indicator, formulated from the question stated in the introduction:

- Health records recognised correctly by Smart Paper Technology Engine.
Ideally, this indicator should reach 100% - this would mean that there were no discrepancies or errors when comparing the data on the documents scanned and data available in the Smart Paper Technology digital database.

For this study, we used materials and information available from the three countries (Uganda, Afghanistan, and The Gambia) where Shifo has conducted pilot studies of the child health component of its solutions. Data was reviewed by two Shifo administrators, who checked for any discrepancies between hand-marked and digital data.

In each country, we used the most recently collected data, at the point of this study, this was data collected in June, July, and August 2017. Data for The Gambia was only available from July and August due to the more recent implementation of Shifo’s solution in the country. In Afghanistan’s case, the dates were slightly different due to the country’s use of the Hijri Shamsi calendar. Exact dates for each evaluation are listed in the table headings in the results section.

In Afghanistan and The Gambia, all visits/documents for the given months were included within our data. They are comprehensive and represent all patients served by Shifo’s child health component in Afghanistan and The Gambia. For Uganda, however, each month utilised a sample size instead of the total number of visits. From each day of June, July, and August in Uganda, documents were randomly selected for evaluation. A sample size for these months in Uganda was taken because of the high number of health facilities using Shifo’s solution in the country. While Afghanistan and The Gambia only utilise the system in a few health facilities, in Uganda it has been integrated in many health facilities throughout the country.

It is important to note that the documents included in this study only relate to child health visits. For our indicator, this means we are measuring data accuracy in relation to provided vaccinations, vitamin supplementation, and deworming per child visit. Check marks/ticks must be identical across handwritten and digital versions of health records to be considered accurate. A comparable Smart Paper database of reproductive, maternal and newborn health has not yet been established in any of the target countries.

3. Results

Data in Smart Paper Forms was compared to data in Shifo’s electronic register to assess how much of the data in Smart Paper Forms was recognised correctly by the Smart Paper Technology Engine. During this study data was collected from 3971 health records in Uganda, 5469 health records in Afghanistan and 5044 health records in The Gambia. For Afghanistan and The Gambia, this represents 100% of children who came to clinics using Shifo’s system during the months of data collection.

The results found from the data collected showed that the accuracy of the scanning and recognition process in Shifo’s health information solutions in all countries was higher than 99% (see tables 1-3). The variations of accuracy, both between and within countries, were lower than one percentage point and showed a stable level of accuracy over the three months of data collection. For the Ugandan sample, the accuracy of data in the electronic register for June, July and August varied between 99.17-99.20%. In Afghanistan, data scanned and processed during the data collection period (M4-M6) had an accuracy of 99.40-99.60%. In The Gambia, data accuracy was 99.80% for the most recent months, the highest result across all three countries. Note that in the Gambian forms, a new function has been introduced that flags to the data verifier if groups of vaccines ticked are not complete, which helps to identify and rectify unrecognised ticks/check marks.
### Table 1: Evaluation results from Uganda

<table>
<thead>
<tr>
<th>Indicator</th>
<th>June 2017</th>
<th>July 2017</th>
<th>August 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of children whose health records were checked from total visits for the month</td>
<td>600</td>
<td>2 750</td>
<td>621</td>
</tr>
<tr>
<td>Percentage of children whose health records were checked from total visits for the month</td>
<td>4 %</td>
<td>26 %</td>
<td>14 %</td>
</tr>
<tr>
<td>Number (percentage) of children whose health records recognised inaccurately</td>
<td>5 (0.80%)</td>
<td>23 (0.83%)</td>
<td>5 (0.80%)</td>
</tr>
<tr>
<td>Percentage of children whose health records recognised accurately</td>
<td>99,2 %</td>
<td>99,17 %</td>
<td>99,2 %</td>
</tr>
</tbody>
</table>

### Table 2: Evaluation results from Afghanistan

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of children whose health records were checked from total visits for the month</td>
<td>1 437</td>
<td>1 775</td>
<td>2 257</td>
</tr>
<tr>
<td>Percentage of children whose health records were checked from total visits for the month</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Number (percentage) of children whose health records recognised inaccurately</td>
<td>7 (0.48%)</td>
<td>11 (0.60%)</td>
<td>10 (0.40%)</td>
</tr>
<tr>
<td>Percentage of children whose health records recognised accurately</td>
<td>99,52 %</td>
<td>99,4 %</td>
<td>99,6 %</td>
</tr>
</tbody>
</table>

### Table 3: Evaluation results from The Gambia

<table>
<thead>
<tr>
<th>Indicator</th>
<th>July 2017</th>
<th>August 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of children whose health records were checked from total visits for the month</td>
<td>1 624</td>
<td>3 420</td>
</tr>
<tr>
<td>Percentage of children whose health records were checked from total visits for the month</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Number (percentage) of children whose health records recognised inaccurately</td>
<td>4 (0.20%)</td>
<td>7 (0.20%)</td>
</tr>
<tr>
<td>Percentage of children whose health records recognised accurately</td>
<td>99,8 %</td>
<td>99,8 %</td>
</tr>
</tbody>
</table>
4. Conclusion

The findings of this study showed that digitisation and data processing using Shifo’s Smart Paper Technology Engine has a very high level of accuracy. As The Gambia had the highest accuracy indications, the conclusion is that the new feature introduced in the verification process contributes to better accuracy. Since, this function has been implemented in the Uganda and Afghanistan as well, increase in accuracy in these countries is expected as well.

This study had a few limitations which should be considered. First, the study only addresses the accuracy of the scanning and recognition of the data entered into the system. It does not assess the quality of information entered by health workers at the point of care. Second, data was only collected during three months and cannot show if accuracy varies over a longer period or is affected by seasonal factors. However, continued collection of data is already being undertaken to address this limitation and additional assessments will be made to continue the tracking of the solution’s accuracy. A strength of this study is the fact that data was collected in three countries representing a diverse set of environments, which shows that accuracy is consistently high across different contexts. The results from this study are therefore transferrable to other countries, provided the system used remains the same.

We can conclude from this study that the accuracy of data provided by Shifo’s solution is high and consistent. Together with further studies planned for the data quality series, we aim to prove that Shifo health information data is of high quality and reliable as a base for decision-making and planning within health care systems.
5. References


