*Article*

# Data Quality: A Negotiator between Paper-based and Digital Records in the Pakistan's TB Control Program

**Syed Mustafa Ali[1], Farah Naureen[1], Arif Noor[1], Maged N. Kamel Boulos[2], Javariya Aamir[1], Muhammad Ishaq[1], Naveed Anjum[1], John Ainsworth[3], Aamna Rashid[1], Arman Majidulla[4] and Irum Fatima[1].**

[1]  Mercy Corps, Pakistan
[2]  University of the Highlands and Islands, UK
[3]  University of Manchester, UK
[4]  Interactive Research and Development, Pakistan
*  Correspondence: mustafa_30_84@yahoo.com; Tel.: 00923018550454;

**Abstract：Background：** Increasingly, healthcare organizations are using technology for the efficient management of data. The aim of this study was to compare the data quality of digital records with the quality of the corresponding paper-based records by using data quality assessment framework. **Methodology：** We conducted a desk review of paper-based and digital records over the study duration from April 2016 to July 2016 at six enrolled TB clinics. We input all data fields of the patient treatment (TB01) card into a spreadsheet-based template to undertake a field-to-field comparison of the shared fields between TB01 and digital data. **Findings:** A total of 117 TB01 cards were prepared at six enrolled sites, whereas just 50% of the records (n=59; 59 out of 117 TB01 cards) were digitized. There were 1,239 comparable data fields, out of which 65% (n=803) were correctly matched between paper based and digital records. However, 35% of the data fields (n=436) had anomalies, either in paper-based records or in digital records. 1.9 data quality issues were calculated per digital patient record, whereas it was 2.1 issues per record for paper-based record. Based on the analysis of valid data quality issues, it was found that there were more data quality issues in paper-based records (n=123) than in digital records (n=110). **Conclusion:** There were fewer data quality issues in digital records as compared to the corresponding paper-based records. Greater use of mobile data capture and continued use of the data quality assessment framework can deliver more meaningful information for decision making.

**Keywords:** mHealth; mobile data collection; data quality; data quality assessment framework; Tuberculosis control; developing countries.

## 0. Strengths and Limitations of the Study

- The scope of study included a comprehensive review and comparison of paper-based and digital data to identify quality issues and differences or mismatches between the two formats, and to categorize the identified issues into the classifiable and non-classifiable ones.
- The strength of the present work is its useful in developing a case for implementation agencies for expanding their digital health initiatives, particularly for data collection.
- Due to patients' information confidentiality concerns and provisions (researchers had no access to or contact with the patients), the researchers were unable to categorize non-classifiable issues (those data that would have required contacting the patient to verify them), which can be considered as a limitation of the current study. Nonetheless, we demonstrated the need for putting in place an adequate data quality improvement strategy so that reliability and sanity of healthcare data can be fully achieved.
- With the limited human and other resources in the enrolled clinics, running two systems (paper-based and digital) in parallel during the study period might have caused frustration among clinic staff. Overburdening the data collection workflows of the involved staff might have also been a reason for

the relatively low (50.4%) overall use of MAPPLE in data collection. With sufficient incentives in place and a complete switch to digital format (following any necessary tweaking and optimization of MAPPLE), digital data collection rates should greatly improve in the future.

## 1. Introduction

With an increased adoption of performance indicators for monitoring the efficiency of healthcare delivery systems, the need for high-quality data generation has also increased [1]. Health information management systems are intended to provide the right information to their users through feedback and data sharing, and are designed for facilitating data-driven decisions, policy making and health planning [2].

Improving the quality of healthcare data is beneficial in many ways, such as, in making well informed decisions about service delivery, ensuring patient safety, conducting research, informing patients regarding their illness and care, and measuring effectiveness of the clinical pathways. Sharing data within and across departments or organizations can provide much needed evidence about healthcare community needs [3], offering a reliable summary of the true health status of patients and the community, and guiding policy makers in making healthcare system adjustments as necessary [4]. Similarly, the cornerstone of public health function is to identify healthcare needs, to influence policy development, and to ensure that any needed services are equitably provided [5].

Organizations rely heavily on various data resources for the effective and efficient management of their operational processes. However, the volume and complexity of some data resources can make them susceptible to defects that can reduce data quality [6] and result in higher operational costs [7]. Data quality (DQ) management aims at objectively measuring quality, with particular emphasis on various data quality aspects [9]; therefore, many DQ management approaches exist that utilize different perspective and have been adopted by organizations [10].

In medical and public health communities, documentation is a critical aspect of DQ and quality of care. Complete documentation records the history of the clinical pathway and its outcomes or effectiveness, providing decision support to healthcare providers. Documentation is commonly maintained in paper-based format in low resource settings [11,12]. Previous research has shown that paper-based information systems tend to produce low-quality data and result in limited or less than optimal data use [13]. The quality of care and quality improvement planning are adversely affected in the case of paper-based information systems; for example, illegibility, incompleteness, and poor organization of records are problems often plaguing the paper format [14].

On the other hand, the benefits of maintaining digital records in healthcare, such as, rapid data sharing, reduced paperwork, lower incidence of medical errors and cost savings, have been commonly discussed in the literature [15,16,17]. Furthermore, with proper digital data security and handling provisions implemented in place, the degree of patient data confidentiality and privacy protections obtained with digital records can exceed that afforded by any paper-based system [18].

Many organizations have started using technology for efficient DQ management because of the huge quantities of data that are involved in their operational processes [9]. Among these technologies, mobile health (mHealth) technologies have gained particular attention for digital data capturing in the public health domain [19]. However, in the absence of an adequate DQ improvement strategy, it becomes challenging to translate data into meaningful information for use in making programmatic and strategic decisions [9]. Moreover, a Data Quality Assessment Framework (DQAF) is a vital constituent of an effective DQ improvement strategy [20].

Data quality improvement strategies involve the definition of DQ dimensions and their standards from an applied "fitness-for-use" perspective. The latter determines the ability of a data collection to meet user requirements. Data quality is a context-specific, multidimensional concept, and the DQ standards to be attained for each proposed use within a specific context are defined by the user [21]. Various data quality improvement frameworks are reported in the literature [22-28]; however, a common aspect among them is a trade-off between DQ dimensions and resource constraints, factoring in organizational and data use preferences [29].

Despite efforts in improving the data quality of paper-based records, the overall data quality remains low, especially in the developing countries. In most of the developing countries, data quality defects are because of information system's inability to detect and prevent errors. In addition, these countries do not adopt context-specific data quality measurement as a usual approach. Since, Mercy Corps Pakistan is digitizing data collection and setting up a computerized management information system for its Tuberculosis Control Program, the objective of this study is to use data quality assessment framework to compare data quality in digital records and their corresponding paper-based records.

## 2. .Methods

### 2.1. Sample Description

Supported by the Global Fund, Mercy Corps Pakistan undertook an mHealth initiative in the Public-Private Mix (PPM) model of the TB control program of Pakistan. In PPM model, all registered care providers are providing free treatment and diagnostic services for TB patients. The sample for this study included six healthcare provider sites (clinics) that qualified for inclusion uniquely because paper-based and digital recording systems were managed simultaneously at each of them. The initiative was focused within the limited geographic areas of the intervention districts (Narowal and Chiniot) and represented by six clinics (three in each district).

Mobile Application for Physician-Patient-Lab Efficiency (MAPPLE) was developed using CommCare platform (https://www.commcarehq.org/) and was tested before its actual use. CommCare software is open source code that can work well with Java-enabled phones. It is an extension to the JavaROSA codebases (code.javarosa.org) that is supporting a range of mobile data collection applications in low-income countries [30]. MAPPLE is an mHealth application loaded with TB-related forms that allows users to enter data in the application and share data with remote cloud server (Figure 1). Before the enrollment of healthcare providers, it was agreed that completing both paper and digital records would be their responsibility during the pilot phase. At clinics, paramedic staff was given responsibility, but there was no incentive system for either clinicians or paramedics. Each healthcare provider was given a smartphone with MAPPLE deployed on it during the month of March 2016.
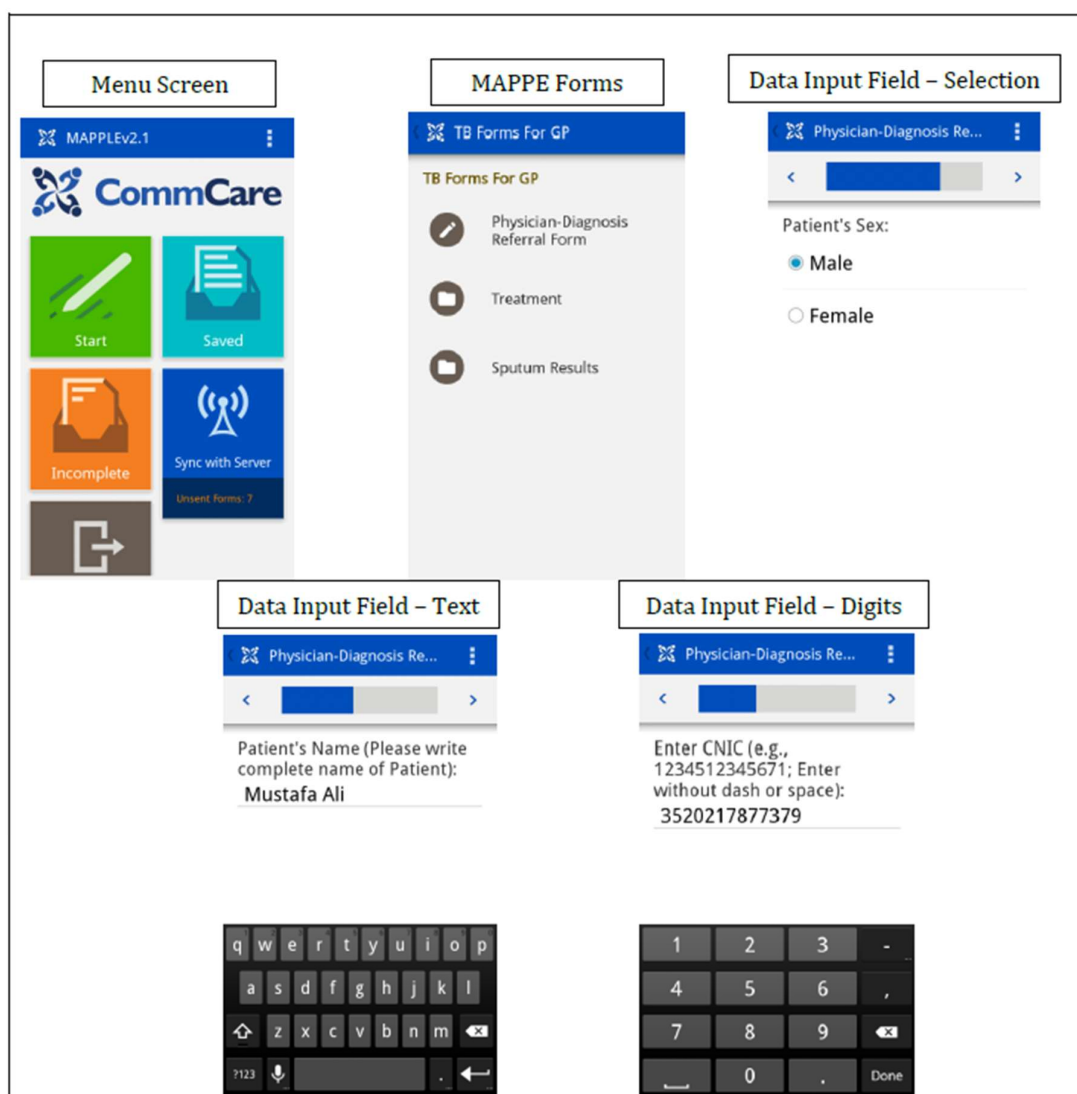
**F**igure 1: Multiple Screenshots taken from Application Workflow.

*2.2. Data Collection*

Paper-based patient treatment cards (TB01 card) prepared during the study period of four months (April 2016 – July 2016) were requested from the six enrolled clinics of the Narowal and Chiniot districts. These enrolled clinics are operated by private and primary healthcare providers, where only one clinician conducts clinical assessment and is helped by support staff, whereas, support staff manages medical stock inventory and patient recording registers. Generally, these healthcare providers are not regulated by health authorities and clinical documentation is also not mandatory.

During the study period, support staff collected both digital data and handwritten data. The copies of TB01 cards were compared with the corresponding digital records, retrieved from the server. TB01 card contains data fields that are representative of the patient's profile, clinical and diagnostic details. TB01 card captures a multi-visit report of a patient's treatment expanded over a period of either six or eight months, depending upon the category of TB patient (CAT I and CAT II). Data fields representing each data category are summarized in table 1.

4

| Category of Data | Data fields | Type of Data | | | |
|---|---|---|---|---|---|
| | | **Number** | **Text** | **Alpha-numeric** | **Selection** |
| Profile | 17 | 5 | 9 | 1 | 2 |
| Clinical | 7 | 3 | 1 | - | 3 |
| Diagnostic | 6 | 4 | 1 | - | 1 |
| **Table 1:** Type and Number of Data Fields on Patient Treatment Card (TB01 card) | | | | | |

## 2.3. Data Quality Assessment Method

Prior to analysis, an approach for logical and comprehensive review was developed and a desk review of the collected paper-based and corresponding digital records of the same service delivery points was conducted. All data fields of the TB01 card were input into a spreadsheet-based template to undertake field-to-field comparison of the shared data fields between TB01 card (paper-based data) and MAPPLE (digital) data (Table 2). Upon culmination of review, non-matching data fields were ordered into classifiable and non-classifiable issues. Classifiable issues were categorized according to the context-specific data quality dimensions, e.g., completeness, accuracy, consistency, understandability and timeliness [37]. Non-classifiable issues were those differences for which correctness or completeness could not be determined without contacting the patient. For example, difference in reported age noted in two formats (digital and paper-based) can only be corrected if patient is contacted for this purpose.

| Category of Data | Comparable Data Fields |
|---|---|
| Profile | 12 |
| Clinical | 6 |
| Diagnostic | 3 |
| Total | **21** |
| **Table 2:** Comparable Data Fields of the Patient Treatment Card (TB01) | |

## 2.4. Data Quality Assessment Framework

Result of field-to-field comparison was summarized using the Data Quality Assessment Framework (QDAF). For the purpose of DQAF development, data quality dimensions were first identified by searching relevant literature [22-28; 31-36] and only those dimensions were selected which were related to the work settings and the field of practice, i.e., Tuberculosis. Utilizing management perspective, the detailed approach for selecting data quality dimensions is reported elsewhere [37]. The DQAF includes DQ dimensions, their operational definitions and method for quantifying data quality issues. It is consistently reported in the literature that data quality uses 'fitness-for-use' perspective of the management – about the work settings and field of practice (TB control in our case) - hence, we preferred to develop our own tool optimized for our own settings, rather than the use of existing tools.

## 2.5. Data Analysis

The operational definitions of the identified data quality dimensions were applied to the data variances for classification purpose.  The non-matching fields between paper-based and digital records were regarded as data quality issue. Each issue was attributed to either paper-based record or digital record, hence called classifiable issue. There were issues which were occurring due to application design modifications; since these issues were emerging because of technology shortcoming or application workflow, which was not aligning clinical workflow, they were excluded from the main dataset. The data quality issues in both paper-based and digital records were recorded against each of the data quality dimension, entered in an Excel sheet. In addition to basic descriptive statistical analyses, a test of proportion was conducted to test the significance of results.

*2.6. Ethical Considerations*

In Pakistan, ethical approval is only required for experimental research involving humans and this study is exempt as it does not qualify as experimental research. However, the study followed all of Mercy Corps' established confidentiality guidelines (https://www.mercycorps.org/research-resources) and was carefully checked by the Monitoring, Evaluation and Learning Unit and Data Controller of Mercy Corps Pakistan.

*2.7. Study Findings*

2.7.1. Data Quality Dimensions:

The DQAF provided a logical basis for reporting the variability in the two datasets (paper-based and digital). The DQ dimensions and their operational definitions, aligned with the field of practice and work settings of the TB control program of Pakistan, are given below (table 3).
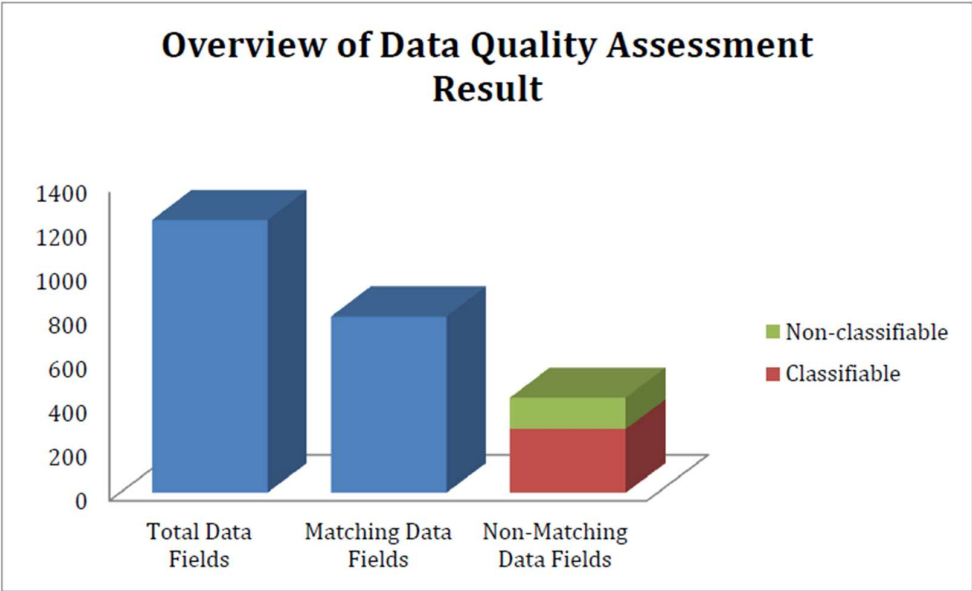
| Dimension | Operational Definition | Adopted from |
|---|---|---|
| Consistency | Representation of data values remains the same in multiple data items in multiple locations | [35] |
| Completeness (data) | Information having all required parts of an entity's description | [31] |
| Accuracy | The measurement or recorded value matches the actual value being captured<br>The degree to which data correctly describes the "real world" object or event being described | [28, 36] |
| Understandability (opposite of Vagueness) | The statement or the term is having clear or specific (unambiguous) meaning | [1] |
| Timeliness | Shared data should be as near real-time as possible. Thus, data should be timely, in that they relate to the present | [33] |
| **Table 3:** Data Quality Dimensions and their Operational Definitions | | |

2.7.2. Comparison of the Paper-based and Digital Records

During the study period, April 2016 – July 2016, a total of 117 TB01 cards were prepared at six enrolled sites, including 68 TB01 cards from three sites in Chiniot district and 49 TB01 cards from three sites in Narowal district. Only 50% of records (n=59; 59 out of 117 TB01 cards) were digitized by paramedics and sent to server, which is a rather low use of the mHealth application (MAPPLE) for the purpose of data collection. TB01 card and MAPPLE had 21 data fields in common, hence the total of 1,239 (n=59 x 21) comparable data fields that were available for analysis.

Out of the 1,239 data fields, 65% (n=803) were found to be correctly matched across paper based and digital records. However, 35% of data fields (n=436) had anomalies either in paper-based records or in digital records. Among the data anomalies, 67% were classifiable (292 out of 436) and 33% were non-classifiable issues (144 out of 436). Non-classifiable issues were the differences in data fields that could not be clearly attributed as an issue neither in the paper-based record nor in the corresponding digital record. Discrepancies in comparable data fields, such as different paper vs. digital values for patient's contact number, national identification number, age, weight, lab serial number could not be settled until feedback from the provider or patient was taken (which was not possible in this study as researchers had no access to the patients in question). These mismatches were therefore categorized as non-classifiable issues. For example, if a patient's age in the paper-based record is 34 and the age of the same patient in the digital record is 42, then this difference was categorized as non-classifiable issue.

Similarly, classifiable issues were those differences in data fields that could be attributed as an issue either in the paper-based record or in the digital record. For example, if in the digital record the age of the patient is given, while in the corresponding paper-based record this field was left empty, then this is considered a paper-based record completeness issue.



Graph 1: Overview of Data Quality Assessment Result.

In an effort to integrate data collection and care delivery processes, within the study period, various design modifications of the data entry forms took place (e.g., making fields 'required', re-organizing questions, adding new form or question to capture missing information), in response to feedback received from application users. Among the classifiable issues, a sub-set of data (n=59) was excluded from the analysis because it had been affected by these design modification activities. Therefore, only valid issues (n=110) of the digital records (DRs) were compared with issues recorded

in the paper-based records (PBR). The distribution of excluded issues in the digital records that occurred due to change in the application design is shown in table 4.

| Data Category | Data Field | # of Issues | Total (n=59) | Data Quality Dimension |
|---|---|---|---|---|
| Patient Profile | Father/Spouse Name | 25 | 37 | Completeness (missing responses) |
| | Type of Referral | 12 | | |
| Clinical | Treatment Start Date | 5 | 22 | |
| | Disease Category | 3 | | |
| | Type of Patient | 12 | | |
| | Disease Site | 2 | | |
| **Table 4:** Quantification and Type of Issues that Occurred due to Design Change | | | | |

Overall, 1.9 DQ issues were calculated per digital patient record, whereas the corresponding figure was 2.1 issues per single paper-based record. Additionally, at the beginning of the study, the number of issues per digital and paper-based records was 1.5 and 2.2 respectively, but these figures later dropped down to 0.7 and 1.4 issues per record respectively by the end of study period. Based on the analysis of valid data quality issues, it was found that there were more DQ issues in the paper-based records (n=123) than there were in the digital records (n=110). A month-by-month comparison of the data showed that April had significantly different entry errors between DR and PBR. In the case of April, errors in the digital records significantly exceeded the digital records. All other months under consideration were not significantly different. The difference between months amongst the digital records showed a significant improvement (p-value=0.0328), while no significant improvements were observed in the case of the paper-based records over time (p-value=0.0629).

| Month | Total comparable records | No. of issues in DR | No. of issues in DR (design concern) | No. of valid issues in DR | No. of issues per DR | No. of issues in PBR | No. of issues per PBR | Test |
|---|---|---|---|---|---|---|---|---|
| | n | x | $x^1$ | $x^v = x - x^1$ | $d = x^v/n$ | y | p=y/n | *p*-value |
| April | 31 | 85 | 40 | 45 | 1.5 | 67 | 2.2 | 0.0287 |
| May | 10 | 50 | 8 | 42 | 4.2 | 34 | 3.4 | 0.3124 |
| June | 10 | 24 | 7 | 17 | 1.7 | 11 | 1.1 | 0.2411 |
| July | 8 | 10 | 4 | 6 | 0.7 | 11 | 1.4 | 0.2138 |
| Total | 59 | 169 | 59 | 110 | 1.9 | 123 | 2.1 | 0.3711 |
| **Table 5:** Monthly Chart of Classifiable Issues in Paper-based Records (PBRs) and Digital Records (DRs) | | | | | | | | |

2.7.3. Analysis of Non-Classifiable Issues

Table 6 lists all 13 data fields where differences were recorded but they were not settled because of patients' confidentiality concerns (no researchers' access to patients). Patient's age was the data field in which most differences were observed, i.e., n= 47. But differences in patient's weight (n=22), among others, were critically important in relation to effective case management, due to the clinical significance of body weight value, its use in patient condition monitoring and its potential to affect certain treatment decisions. Issues with the patient identifier code (n=20) were also of considerable significance.

| Name of Data Fields | April | May | June | July | Total |
|---|---|---|---|---|---|
| Name | 0 | 0 | 2 | 0 | 2 |
| Age | 25 | 12 | 8 | 2 | 47 |
| Weight | 12 | 2 | 6 | 2 | 22 |
| Patient Identifier Code | 3 | 9 | 4 | 4 | 20 |
| National Identity Number | 6 | 3 | 1 | 1 | 11 |
| Address | 1 | 0 | 0 | 1 | 2 |
| Phone Number | 6 | 1 | 1 | 1 | 9 |
| Father/Spouse Name | 0 | 2 | 0 | 0 | 2 |
| Supporter Name | 1 | 0 | 0 | 0 | 1 |
| Type of Referral | 6 | 11 | 3 | 2 | 22 |
| Lab Number | 2 | 0 | 1 | 0 | 3 |
| Disease Site | 0 | 2 | 0 | 0 | 2 |
| Lab Result | 1 | 0 | 0 | 0 | 1 |
| Total | 63 | 42 | 26 | 13 | 144 |
| **Table 6:** Data Field-wise Distribution of the Non-classifiable Issues | | | | | |

2.7.4. Analysis of Classifiable Issues

All valid classifiable issues (excluding those issues that occurred because of the aforementioned design modifications) were further categorized according to data quality dimensions. Overall, there were more completeness issues (n=148; 63.5%), followed by timeliness (n=44; 19%), accuracy (n=30; 13%), understandability (n=10; 4%) and consistency issues (n=1; 0.5%) in the set of valid classifiable issues.
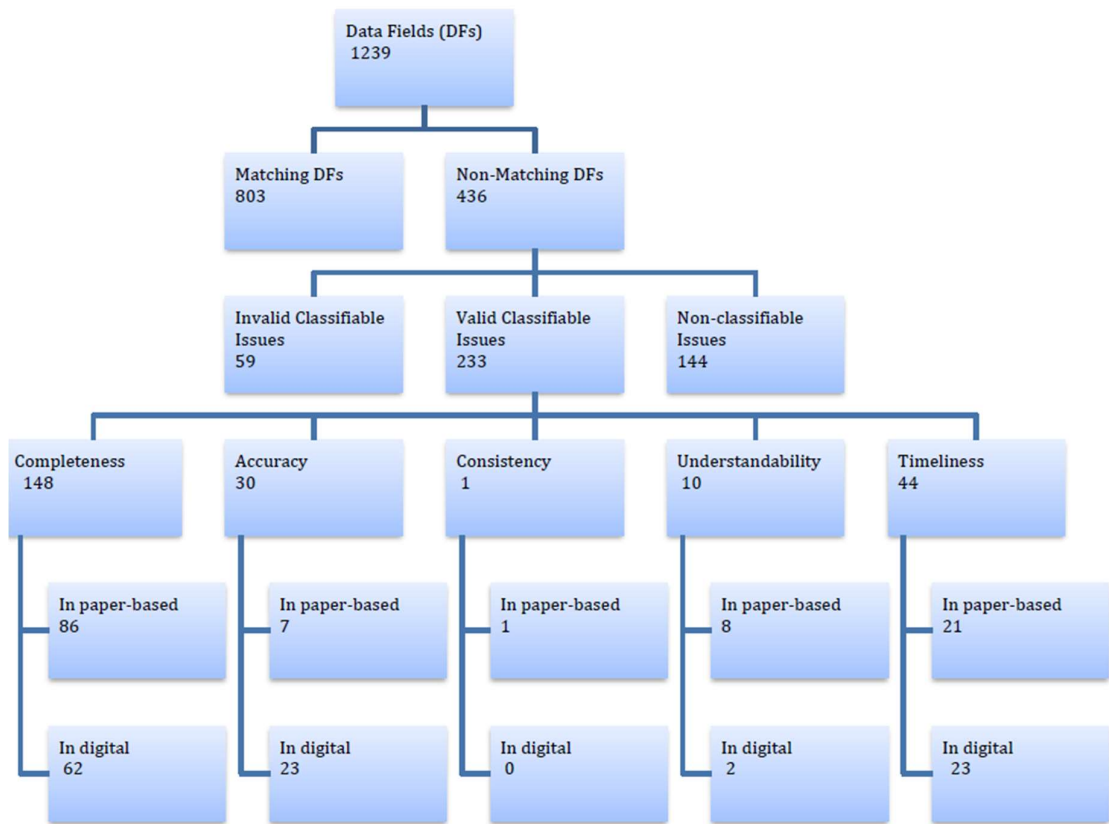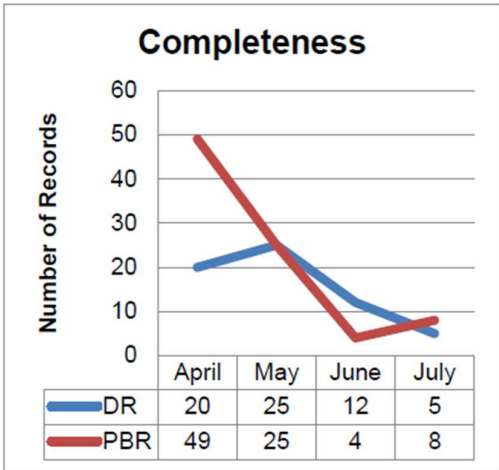
Figure 2: Data Quality Classification Tree.

The detailed findings of the data quality assessment exercise are presented below categorized by data quality dimension.

Classifier 1: Completeness

An operational definition of Completeness is "information having all required parts of an entity's description" [31].
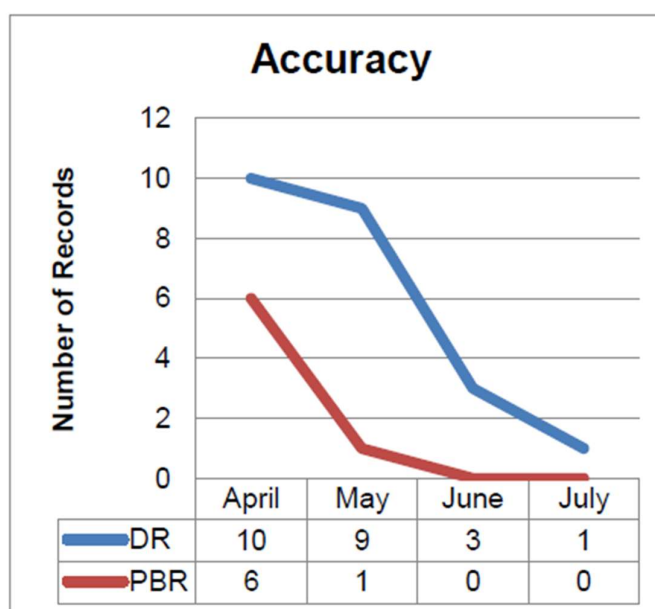


| | April | May | June | July |
|---|---|---|---|---|
| DR | 20 | 25 | 12 | 5 |
| PBR | 49 | 25 | 4 | 8 |

Data completeness issues were found in both datasets; however, there were more such issues in the paper-based medical records, i.e., 58% of all observed data completeness issues. Upon further analysis, if was found that Patient Name (n=9) and Address (n=14), and Treatment Supporter Name

(n=11) and Address (n=12) were the digital data fields that showed more issues of completeness. In the paper-based records, the top completeness issues were in patient's address (n=10), type of referral (n=11) and laboratory examination date (n=7). Therefore, it can be said that most of the encountered data completeness issues were in the patient profile data types that allowed free text input and hence were more prone to errors. However, there were relatively less observed completeness issues in clinical and diagnostics data types, except for laboratory examination date.

Classifier 2: Accuracy

Applying understanding of the field of practice (Tuberculosis treatment) and work settings, accuracy can be defined as *"the degree to which data correctly describe the "real world" object or event being described"* [28].



Accuracy is one of the key data quality dimensions that helps data user in build trust in data representativeness. Data-field-level analysis showed that most of such issues were found in the digital records. Out of total 30 observed accuracy issues, 77% (n=23) were found in the digital records, and most of these issues were in Patient Identifier Code (n=12) and National Identity Card Number (n=6).
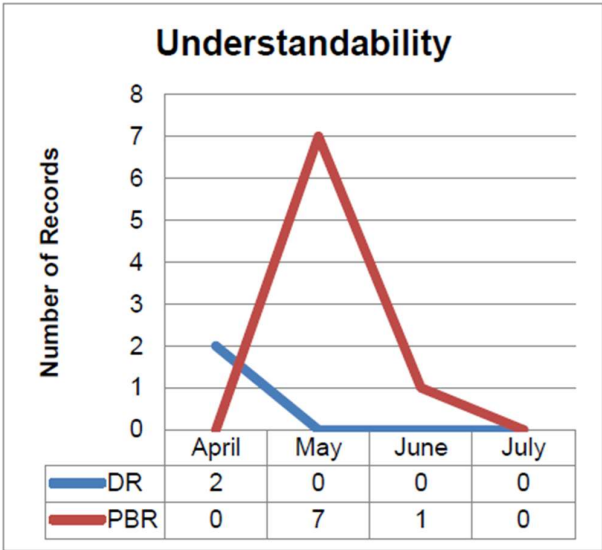
Classifier 3: Consistency

By consistency we mean that the "representation of data values remains the same in multiple data items in multiple locations" [35].

Consistency was the least reported issue type in our set, with only one issue found in the paper-based records.
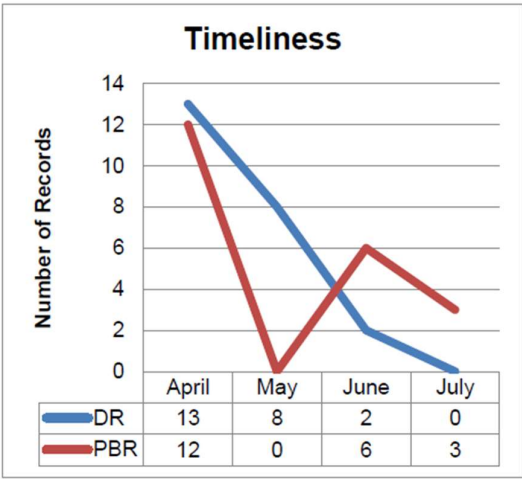
Classifier 4: Understandability

Utilizing the "fitness-for-use" perspective, understandability can be defined as *"the statement or the term that has clear or specific meaning"* [1].



Under understandability, the findings of this data quality assessment exercise can be mainly linked to one of the commonest issues associated with paper-based records, namely illegibility of handwriting. There were a total of 10 understandability issues in our set, and most of them were spotted in the paper-based records (n=8; 80%).

Classifier 5: Timeliness

Under the MAPPLE mHealth initiative, timeliness means that "shared data should be as near real-time as possible. Thus, data should be timely, in that it relates to the present" [33].

**Timeliness**

| | April | May | June | July |
|---|---|---|---|---|
| DR | 13 | 8 | 2 | 0 |
| PBR | 12 | 0 | 6 | 3 |

Though the general principles of informatics encourage the integration of application and clinical workflows, technology use also ensures the timeliness of data recording and reporting. However, in our studied set, there were slightly more timeliness issues observed in digital records than in paper records (DR=23; PBR=21; difference=2), which is a clear indication of the weak integration between workflow processes. Besides the importance of integrating workflows, treatment start date and lab exam date are also of critical importance for achieving the desired health outcomes monitoring of treatment timeline. It was observed that all of these issues (n=44), either in paper-based records or in digital records, were in those data fields storing treatment start date and follow-up evaluation dates.

## 3. Discussion

Global evidence identifies high data quality as a necessary condition for the delivery of quality healthcare [38]. In the developing countries, health information systems are needed to tackle the growing public health concerns, as current paper-based documentation systems are becoming increasingly inadequate [39]. Therefore, mHealth technology is being implemented in the public health settings of the developing countries.

This study looked at the paper-based records and their corresponding digital records at the six points or locales of TB care that have started using a mobile data collection application (MAPPLE) from March 2016. As a theoretical framework is helpful in addressing data variability issues [29], we developed a data quality assessment framework to assess data quality. According to the study's findings, digital records have generated better data quality in the first quarter of their implementation. On the other hand, despite years of staff's practice in maintaining the paper-based patient record, our assessment results showed relatively poor data quality associated with handwritten paper forms. Additionally, factors such as unregulated and non-standardized practices in the developing countries, and non-incentivized data collection in private healthcare settings highlight the significance and future implications of these findings.

Currently, in the Public-Private Mix model of TB care delivery, there are multiple stakeholders representing different levels of the management within an organization and across different organizations. The complexity in the management structure demands high- level of collaborative relationship between different management units [29]. However, the problem of management complexity can be addressed if different organizations have similar level of direct control over the data they generate during their normal care and management procedures [40]. Hence, all

stakeholders get an equal opportunity for the data quality review. Therefore, organizations will start producing high quality data by strategizing the use of data quality assessment framework.

Data quality issues were found in all three data types: patient profile, clinical and diagnostic data. Issues in the clinical variables are of critical importance [18]. As part of data quality improvement strategy, there should be a mechanism to flag disparities in the clinically important data fields [41]. Errors in clinical practice are sometimes attributed to medical documentation errors in paper-based records [42], but digital records, when not properly designed and implemented, can equally suffer from data inaccuracies leading to medical errors [18]. Furthermore, it is critically important to receive complete and correct patient information, and it is achievable if mHealth technology is fully exploited beyond the mere basic functions of digital data collection, storage, retrieval and sharing [41, 43].

### 3.1. User Adoption and Acceptance Issues of Digital Data Collection

Though there was some improvement in the data quality of digital records over the study period of four months, there was also a gradual decrease in the use of MAPPLE (mobile application). This might be due to frequent application design modifications and non-incentivized data collection. The current use of the MAPPLE, used primarily for data collection at the six study sites, is inconsistent and without any supportive supervision or management's active role in ensuring the regular use of application. Additionally, no reward mechanism was introduced to encourage application use for the purpose of data collection.

It has been observed that data collection, digitization and aggregation are increasingly difficult tasks in developing countries [44] due to the lack of incentive programs [45]. Additionally, application design considerations should include making all required functions available on the user's device in a highly usable and intuitive fashion [46]. Applications should be designed with full user involvement from the early design stages and throughout the application's lifecycle, including its regular maintenance and updates. Applications should seamlessly integrate with existing clinical workflows, improving rather than overburdening them, and taking into consideration the already high work and cognitive loads of most healthcare professionals today [47]. Free text input should be kept to a minimum in digital forms (also to avoid errors), and clear and comprehensive choices offered, instead for users to select from them. Integrity and validation checks should be built into digital forms. Other strategies for minimizing user input, reducing errors and improving acceptance include cross-linking relevant databases to 'autocomplete' certain fields where applicable, based on values entered in other fields.

Improving data quality is task-dependent and includes aligning data collection processes, operationalizing quality improvement strategy, and building capacity for those responsible for data entry and review. Therefore, with an application like MAPPLE, there is a wide range of organizational and system-specific factors that may affect the adoption of healthcare information technology [48].

### 3.2. Novel Contribution, Replicability and Generalizability of the Work beyond the Six Study Locales

The novel contribution made by this study concerns our model of using an assessment framework that is inclusive of the management perspective and is more relevant to local work settings and field of practice. We believe that a meaningful assessment would not have been possible had we opted to use existing frameworks (generic or developed for other contexts), as only the local data users can conceptualize and contextualize data quality [21,29]. Long ago, it has been identified that the definition of data quality varies between users, locales and contexts, which makes the data quality concept multi-dimensional and complicated [49,50]. Considering this, a similar approach was also used elsewhere [1, 51], that included the perspective of data users, but our study offers a detailed example and model to follow, particularly in the context of developing countries (most existing literature focuses on examples for affluent countries and their settings).

Though the study included six participating primary healthcare clinics, it is observed that across the country the characteristics of clinics and their clinical and data management practices remain nearly same [52]. The private healthcare system remains largely un-regulated due to lack of interest of public health authorities [53]. This provides non-governmental organizations (NGOs) with an opportunity to bridge the gap between service need and service provision [53]. Since the Public-Private Mix model is working well in 75 districts of Pakistan, our framework and findings can be replicated in other districts of the country (and other countries sharing our settings) when digitization and data quality improvement plans are rolled out in those places.

*3.3. Research Implications*

This study included a review of patient records in paper and digital formats, and concluded that, in the studied set of records, digital data were of moderately better quality compared to data from the corresponding paper-based records. For significant and sustained improvement in data quality, the study emphasized the need for putting a comprehensive DQ improvement strategy in place and making incentive program a significant part of it. Our Data Quality Assessment Framework (DQAF) offered a comprehensive and unified approach to staff who managed the data. Based on this research, DQAF is proposed to be included in the current data quality improvement strategy of the organization, to ensure that reliable data are available for effective decision making. The research team also plans to conduct DQ assessment of the same points of TB care after the inclusion of DQAF's use in routine data review practices.

**4. Conclusions**

The Data Quality Assessment Framework (DQAF) is evidence based and draws its concepts and constructs upon reliable literature sources. The framework was applied to two datasets – paper-based records and digital records – and utilized to elucidate the usability of DQAF and understand the data quality dimensions that used the "fitness-for use" perspective. In relevance to work settings and field of practice, i.e., TB treatment, the DQAF used the dimensions of completeness, accuracy, consistency, understandability and timeliness for inter-system analysis and review.

Using DQAF, it was found that the overall quality of digital records is moderately better than the quality of paper-based records.   Therefore, in addition to the presence of data quality improvement strategy, the use of DQAF should also be introduced as routine practice. Likewise, considering the inherent ability of the technology in improving data quality, design modifications and workflows optimization and integration should also be considered essential for the adoption of mHealth technology. Consequently, strengthening of the information management system would help organizations in building trust in data, making evidence-based and informed decisions about health policy and practice.

**References**

1.  St-Maurice, J., Burns, C.M.: A method for developing data quality measures and metrics for primary health care. In *Proceedings of International Symposium on Human Factors and Ergonomics in Health Care: Improving the Outcomes* (2016). DOI: 10.13140/RG.2.1.2972.5048

2.  Ndabarora, E., Chipps, J.A., Uys, L.: Systematic review of health data quality management and best practices at community and district levels in LMIC. *Inf Dev*. 1-18 (2013). DOI: 10.1177/0266666913477430

3.  Kerr, K.A., Norris, T., Stockdale, R.: The strategic management of data quality in healthcare. *Health Inform J*. 14(4), 259-266 (2008).

4.  Haux, R.: Health information systems past, present, future. *Int J Med Inform*. 75(3-4), 268-281 (2006).

5.  Walker, R.: Health information and public health. *Health Inf Manag J*.  37, 4-5 (2008).

6.  Even, A., Shankaranarayanan, G.: Dual assessment of data quality in customer databases. *ACM J Data Inf Qual*. 1(3), 1-29 (2009).

7.  Redman, T.C.: Data: An unfolding quality disaster. *DM Review*. 14(8), 21-23 (2004).

8.  Cappiello, C., Francalanci, C., Pernici, B.: Data quality assessment from the user's perspective.  In *Proceedings of the 2004 International Workshop on Information Quality in Information Systems*. 68-73. DOI: 10.1145/1012453.1012465

9.  Madnick, S.E., Wang, R.Y., Lee, Y.W., Zhu, H.: Overview and framework for data and information quality research. *ACM J Data Inf Qual*. 1(1), article 2, 1-22 (2009).

10. Glowalla, P., Sunyaev, A.: Process-driven data quality management: A critical review on the application of process modeling languages. *ACM J Data Inf Qual*. 5(1-2), article 7, 1-30 (2014).

11. Agyeman-Duah, J.N.A., Theurer, A., Munthali, C., Alide, N., Neuhann, F.: Understanding the barriers to setting up a healthcare quality improvement process in resource-limited settings: a situational analysis at the medical department of Kamuzu Central Hospital in Lilongwe, Malawi. *BMC Health Serv Res*. 14:1 (2014). DOI:10.1186/1472-6963-14-1

12. Ogunsola, O.O., Aburogbola, F., Olajide, O., Ladi-Akinyemi, B.: Clinical documentation and doctor: Is it a challenge in HIV care? Experience of four new comprehensive HIV sites in Oyo State, Nigeria. *Adv Trop Med Pub Health Int*. 5(3): 77-89 (2015).

13. Lium, J.T., Tjora, A., Faxvaag, A.: No paper, but the same routines: a qualitative exploration of experiences in two Norwegian hospitals deprived of the paper based medical records. *BMC Med Inform Decis Mak*. 8:2 (2008).

14. Pourasghar, F., Malekafzali, H., Koch, S. Fors, U.: Factors influencing the quality of medical documentation when a paper-based medical records system is replaced with an electronic medical records system: an Iranian case study. *Int J Technol Assess Health Care*. 24(4), 445-451 (2008).

15. Wager, K.A., Lee, F.W., White, A.W., Ward, D.M., Ornstein, S.M.: Impact of an electronic medical record system on community-based primary care practices. *J Am Board Fam Pract*. 13(5), 338-348 (2000).

16. Makoul, G., Curry, R.H., Tang, P.C.: The use of electronic medical records: communication patterns in outpatient encounters. *J Am Med Inf Assoc*.  8(6), 610-615 (2001).

17. Ammenwerth, E., Eichstadter, R., Haux, R., Pohl, U., Rebel, S., Ziegler, S.: A randomized evaluation of a computer-based nursing documentation system. *Method Inform Med*.  40, 61-68 (2001).

18. Ozair, F.F., Jamshed, N., Sharma, A., Aggarwal, P.: Ethical issues in electronic health records: A general overview. *Perspect Clin Res*. 6(2), 73–76 (2015).

19. 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 Global Health*. 3(2), 1-3 (2013).

20. Sebastian-Coleman, L.: Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework. Waltham, USA, Morgan Kaufmann Elsevier (2013).

21. Pringle, M., Wilson, T., Grol, R.: Measuring 'goodness' in individuals and healthcare systems. *BMJ*. 325 (7366), 704-707 (2002).

22. Tayi, G., Ballou, D.: Examining data quality. *Communications of the ACM*. 41(2), 54-57 (1998).

23. Redman, T.C.: Data Quality: The Field Guide. Boston, Digital Press (2001).

24. Pipino, L., Lee, Y., Wang, R.: Data quality assessment. *Communications of the ACM- Supporting Community and Building Social Capital*. 45 (4), 211-218 (2002).

25. Scannapieco, M., Catarci, T.: Data quality under the computer science perspective. *Int J Architectural Comput*.  1-12 (2002).

26. Maydanchik, A.: Data Quality Assessment. Brandly Beach, NJ, Technics Publications (2007).

27. Batini, C., Cappiello, C., Francalanci, C., Maurino, A.: Methodologies for data quality assessment and improvement. *ACM Comput Surv.* 41(3), 1-52 (2009).

28. DAMA.: *The Six Primary Dimensions for Data Quality Assessment: Defining Data Quality Dimensions*. DAMA UK Working Group on Data Quality Dimensions. https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQ DimensionsWhitePaperR37.pdf (2013). Accessed on 5th May 2016

29. Kahn, M.G., Raebel, M.A., Glanz, J.M., Riedlinger, K., Steiner, J.F.: A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care*. 50, S21-9 (2012).

30. Mhila, G., DeRenzi, B., Mushi, C., Wakabi, T., Steele, M., Roos, D., Dhadialla, P., Sims, C., Jackson, J., Lesh, N. Using mobile applications for community-based social support for chronic patients. 6th Health Informatics in Africa Conference (*HELIN*A), Pan Africa Health Informatics Association, Abidjan (2009).

31. Bovee, M., Srivastava, R., Mak, B.: A conceptual framework and belief-function approach to assessing overall information quality. *Int J Intell Syst*. 18(1), 51-74 (2001).

32. WHO.: Improving Data Quality: A Guide for Developing Countries. World Health Organization – Regional Office for the Western Pacific.: http://www.wpro.who.int/publications/docs/Improving_Data_Quality.pdf (2003). Accessed 22nd June 2016

33. Orfanidis, L., Bamidis, P.D., Eaglestone, B.: Data quality issues in Electronic Health Records: an adaptation framework for the Greek health system. *Health Inform J*. 10(1), 23-36 (2004).

34. Canadian Institute for health information.: The CIHI Data Quality Framework. Ottawa, CIHI). https://www.cihi.ca/en/data_quality_framework_2009_en.pdf. (2009)[Accessed on 5th May 2016]

35. Almutiry, O., Wills, G., Alwabel, A., Crowder, R., Walters, R.: Toward a framework for data quality in cloud-based health information system. *International Conference on Information Society (i-Society) 24-26 June 2013*. http://ieeexplore.ieee.org/document/6636362/. Accessed on 3rd July 2016

36. Pratiwi, A.S., Anawar, S.: A theoretical framework of data quality in participatory sensing: A case of mHealth. *Jurnal Teknologi*. 77(18), 137-146 (2015).

37. Ali, S.M., Anjum, N., Boulos, M.N.K., Ishaq, M., Aamir, J., Haider, G.R.: Measuring management's perspective of data quality in Pakistan's tuberculosis control programme: a test-based approach to identify data quality dimensions. BMC Research Notes. 11: 40 (2018).

38. Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S.C., Shekelle, P.G.: Systematic review: impact of health information technology on quality, efficacy, and costs of medical care. *Ann Intern Med*. 144(10), 742-52 (2006).

39. Kalogriopoulos, N.A., Baran, J., Nimunkar, A.J., Webster, J.G.: Electronic medical record systems for developing countries: review. In *Proceedings of Annual International Conference of IEEE in Engineering in Medicine and Biology Society*. Mineapolis, USA, 3-6 September 2009, 1730-1733 (2009).

40. Divorski, S., Scheirer, M.A.: Improving data quality for performance measures: results from a GAO study of verification and validation. *Eval Program Plann.* 24(1),83-94 (2001).

41. Chen, H., Hailey, D., Wang, N., Yu, P.: A review of data quality assessment methods for public health information systems. *Int J Environ Res Public Health*. 11(5), 5170-5207 (2014).

42. Tsai, J., Bond, G.: A comparison of electronic records to paper records in mental health centers. *Int J Quali Health Care*. 20(2):136-143 (2008).

43. Ward, M., Brandsema, P., van Straten, E., Bosman, A.: Electronic reporting improves timeliness and completeness of infectious disease notification, the Netherlands, 2003. *Euro Surveillance* 10(1), 27-30 (2005).

44. Shovlin, A., Ghen, M., Simpson, P., Mehta, K.: Challenges facing medical data digitization in low-resource contexts. In: Global Humanitarian Technology Conference, IEEE 2013, San Jose, 365-371 (2013).

45. Bram, J.T., Warwick-Clark, B., Obeysekare, E., Mehta, K.: Utilization and monetization of healthcare data in developing countries. *Big Data*, 3(2), 59-66 (2015).

46. Duhm, J., Fleischmann, R., Schmidt, S., Hupperts, H., Brandt, S.A.: Mobile electronic medical records promote workflow: physicians' perspective from a survey. *JMIR Mhealth Uhealth,* 4(2): e70 (2016).

47. Despont-Gros, C., Rutschmann, O., Geissbuhler, A., Lovis, C.: Acceptance and cognitive load in a clinical setting of a novel device allowing natural real-time data acquisition. *Int J Med Inform.* 76(11-12):850-855 (2007).

48. Bassi, J., Lau, F., Lesperance, M.: Perceived impact of electronic medical records in physician office practices: a review of survey-based research. *Interact J Med Res*. 1(2): e3 (2012).

49. Wand, Y., Wang, R.Y.: Anchoring data quality dimensions in ontological foundations. *Communications of the ACM.* 39(11), 86-95 (1996).

50. Wang, R.Y., Strong, D.M.: Beyond accuracy: What data quality means to data consumers. *J Manage Inform Syst*. 12(4): 5-33 (1996).

51. National Health Service.: Executive Summary of the First National Data Quality Review. Quality Information Committee of the National Health Service. Available from: https://www.england.nhs.uk/wp-content/uploads/2013/04/1ndqr-exec-sum.pdf (2013). Assessed on 25th February 2017

52. Nishtar, S.: The Gateway Paper on Health System in Pakistan – a Way Forward. Pakistan's Health Policy Forum and Heartfile, Islamabad, Pakistan. Available from: http://www.heartfile.org/pdf/phpf-GWP.pdf (2006). Accessed on 25th February 2017

53. Shaikh, B.T.: Private sector in health care delivery: A reality and a challenge in Pakistan.  *J Ayub Med Coll.* 27(2): 496-498 (2015).