# Improving Data Quality in a Resource Constraint Public Health Organization in Nigeria with Divide and Conquer and Lot Quality Assurance Sampling Approach

Stephen Boerwhoen Dapiap, Babatunde Adeshina Adelekan, Ahmad Tijjani Aliyu

Department of Strategic Information,
Institute of Human Virology,
Abuja, FCT, Nigeria

## ABSTRACT
Inconsistency or unstable implementation policies in electronic records management systems are likely to create discrepancy in data which may distorts facts as quality of information is compromised. This leads to taking misleading decisions and actions which may be life threatening circumstances in health settings, business losses, misplacement of priorities and wrong interventions for development. Identifying the causes of data discrepancies should be a priority in driving efforts to improve quality at different levels of data ecosystem. This paper investigated the implication of frequent changes in electronic health records management system implementation policy on data quality in an organization supporting health facilities providing HIV/AIDS services across twelve states in Nigeria through the application of divide and conquers and lot quality assurance sampling methods. Large data discrepancies were discovered using the combined methods and there was tremendous data quality improvement six-month after the intrinsic and contextual data quality validation. The study concluded that frequent changes in electronic data management systems are likely to breed distortions in data quality that may greatly affect effective delivery of the most needed quality services.

## General Terms
Data Analytics, Data Quality Assurance, Data Quality Improvement.

## Keywords
Electronic Medical Records, Public Health, Data Quality, Policy Changes.

## 1. INTRODUCTION
Data is one of the most valuable organizational resource or asset (Teradata, 2013; Warme, 2014; Herring, 2016) and as stated by Anita Chung of IBM, "Organizations of all sizes need to rely on their core assets" (IBM, 2016; Horthonworks, 2016). This core asset Chung refers to is data. Andrew Lo (MIT, 2016) once said "For most companies, their data is their single biggest asset". Data has a generic definition of being unorganized or raw facts have different types depending on what generates them. Thus data is more of domain specific and can be considered as a product (Wang and Strong, 1996). There is business data, scientific data, health data, and so forth. The correctness or otherwise of data determines the flawlessness or otherwise of information. Data fit for intended use demands the minimum necessary quality.The issue of data quality is as old as data itself (Sadiq et al., 2011). Even though over two decades there have been research in data and information quality (Shankaranarayanan and Blake, 2017), data quality issues are still very much plaguing organisations especially in low and medium income countries. However, there is additional focus on quality of usage and context rather than only on the measurement and assessment of data quality content. Data quality has profitable or lost impact on strategic decision making, customer satisfaction, earned public confidence and organizational trust according to Sadiq et al. (2011). Numerous business initiatives have been delayed or even cancelled, citing poor-quality data as the main concern (Gee and Helfert,2013) and thus the implication that poor data quality can have substantial social and economic impacts (Wang and Strong,1996). Due to the usefulness attached to data, data is described as a form of capital and placed at the same level as financial capital as it generates new digital products and services (MIT, 2016). Firms are therefore, advised to regard data as raw materials and they are at risk if data is taken for granted.

Health institutions especially have thrived over the years to manage patients' records in the best possible ways for efficient patient management and monitoring. In the developed economies where there are available resources, robust electronic health record systems are available even in local clinics but that is not the case in middle and low income countries (Piette et al., 2012). Many resource-limited health settings in middle and low income countries are still struggling to embrace electronic health record system and if decided to adopt one, yet to fully adapt to it. Institute of Human Virology Nigeria (IHVN) is an International non-governmental organization supporting people living with HIV/AIDS and people affected by HIV/AIDS, Malaria prevention and treatment and Multi-drug resistant tuberculosis diagnosis, care and treatment. IHVN's main intervention programme from inception was the US President's Emergency Plans For HIV/AIDS relief (PEPFAR) started in 2004 with seven (7) facilities and by 2015 has grown to over 300 facilities offering HIV services alone. The data management for patient management and monitoring in this program was initially using simple excel spreadsheet and later Microsoft access database management system. As the number of facilities and patients increased the data management system was changed to CareWare - a patient electronic record management system (http://www.jprog.com/wiki/All-CAREWare-documentation.ashx). The introduction of CareWare to manage the patients' records came with different data quality challenges such as incomplete data, missing data, data transcription errors, duplicated entries, data retrieval difficulties and so forth, emanating from either data entry errors or primary data sources (filling of data capturing tools e.g. forms).

With continuous personnel training, data verification and validation, data gaps and other errors were gradually managed but not completely eliminated. The verification and validation was the examining and enhancing of the actual data (Peer et al. 2014). However, as of 2011 CareWare was deployed to only 14 of the 84 facilities initially intended to implement the electronic medical records system. Due to some financial requirements to implement CareWare at all the 84 facilities including the technical consultants who were off country, the management of IHVN decided to introduce another patient records management system – Teleform which was partly paper-based and partly electronic. The introduction of teleform system might have been a way of addressing scalability problem to cover all the facilities with electronic medical records system with reduced operational costs (Ajami and Bagheri-Tadi, 2013).

Teleform was deployed to all the facilities and a few months after its implementation, many gains from CareWare-implemented facilities were suddenly lost. There were serious data quality issues such as patients' records depletion (shortfall), inconsistent and incomplete patients' identifiers, missing data values, irreconcilable monthly data summaries with aggregate data directly reported from hard copy summary tools e.g. registers at the facilities, duplicated or multiple same patient's records and delays and so forth. As the number of facilities and patients keep on increasing, the data quality issues were increasing almost geometrically. The tendency for size increase in data was also confirmed by IBM's Anita Chung who said that "data volume is doubling every two years for the average organization, and a lot of that data must be managed for years" (IBM, 2016). Teleform system (https://en.wikipedia.org/wiki/TeleForm) has the size limitation as it uses Microsoft access as the backend in addition to the data quality issues. To address the issues inherent with teleform therefrom, couple with its partial electronic nature, an open source electronic health management records system- Open Medical Records System (Open MRS) was suggested, introduced, tested at few facilities and later deployed to over two-third of the organization's facilities. One of the objectives of this system was to have a rugged, robust and stable database for efficient, reliable and effective storage and retrieval, as in the words of Richard Wozniak, "database is the foundation of all information management" (IBM, 2016). The initial deployment was at the high volume facilities and the process of populating the databases was migrating data (Elamparithi and Anuratha, 2015) from the existing facilities' teleform data sets. These in effect made the databases inherited all the data quality issues (Lin and Chen, 2000) inherent from the teleform data sets.

With the adoption of Open MRS as the organization's patient electronic medical record system, measures were initiated to verify, validate and clean the various facilities databases. Data quality review was advocated for by (Peer et al., 2014). Peer et al. (2014) described data quality review as "a process whereby data and associated files are assessed and required actions are taken to ensure files are independently understandable for informed reuse". Improving and maintaining data quality initiative included the onsite data verification and validation: intrinsic, contextual and representation of data quality (Chang and Strong, 2013). This work aims at demonstrating and sharing the approach to improving patient level data collection, data quality monitoring and corrections and other data discrepancies witnessed from the daily health services deliveries at the HIV clinics with a view to having a jewel-like, perfect databases at all the facilities and to rebuild all stakeholders confidence in the organisation's client record management system. In this work, we look intrinsically at the current system data proportion of errors which were inherited by the new system - Open MRS, from the old system –teleform. The errors from the facilities databases were profiled and represented in Table 1.

# 2. METHODS

## 2.1 Identification of Facilities and Drafting of Data verification and validation Protocol

Facilities were identified using the list where the new electronic medical record (EMR) system was deployed. Data verification and validation process protocol was developed followed by orientation on the protocol for staff to do the verification and validation exercise. The aim of the protocol was to help with the standard of logically carrying out the exercise. The defined logical structure was to ensure total quality dimensions inclusion, for instance, capturing all clients ever registered (or transferred in) at every facility, all client visits, and so forth. To achieve the stated aim, all gaps that might exist at the site level as a result of deviation from the institution's standard data management practices were to be identified. The implication was to validate every patient record registered at a site from the source documents against that in the EMR database and update any missing patient record. But achieving this for every patient for high volume sites was herculean and thus had to use a combination of methods - lot quality assurance sampling (LQAS) (Talc., 2003; Hund, 2014; USAID and NUMAT, 2010) and divide and conquer (DAC) (Blelloch, 2011) to select the patients' folders at random and divide the contents base on thematic areas such as pharmacy, lab, and so forth chronologically.

## 2.2 The LQAS and DAC Approach for Validation at the Facilities

In order to identify the gaps at a facility database, the divide and conquer (DAC) and lot quality assurance sampling methods were used to sample the patients' folders. First, the DAC approach was used to divide the facilities with electronic medical record (EMR) system among the teams and number of folders into the enrolment years (EYs) depending on when a facility started the program of enrolling patients into care. For instance, a facility that started the program in 2005, has EYs of n= 12 with indexes of 2005… 2016. The LQAS was therefrom employed to sample folders from EYs. But before the sampling, the summary source document, which is the enrolment register, was used to ascertain that every enrolee in the register was also in the EMR database and vice versa. The enrolment register is a summary tool of all pre-treatment and on-treatment patients registered sequentially with PEPFAR identifiers in yearly cohorts. Profiled gaps identified in either way were documented, for example, records in the register not in the database were created in the database, and if otherwise, they were documented in the register.

Figure 2 illustrates the first schedule of the exercise of authenticating the registered patients at a facility. The source documents used were the summary paper registers of patients' enrolment in care and later initiated on antiretroviral therapy (ART) based on meeting the eligibility criteria – pre-ART (enrolment in care) and ART (initiation on ART) registers. The steps entailed: printing/extracting all unique records of patients registered at the facility by cohorts from EMR database indication with basic identifiers and demographics such as Pepfar identifier (unique), hospital number (unique-facility specific), date of birth, sex, names, enrolment into care

date, and ART initiation date. These characteristics were then validated against the pre-ART register and ART register. If patient was on the register and missing from EMR, the patient record was created into the EMR database and all records of services ever received by the patient were entered as well. The reverse was effected, if on the other hand, a patient was in EMR and missing on any of the registers by retrieving the patient's folder to validate, then if authentic, the patient's record included in the right cohort in both registers if the patient was on treatment, otherwise only in the pre -ART register.

The process of verification and authentication of registered or omitted supposedly registered patients was easier with the DAC approach as the unique identifiers of patients were categorized based on enrolment date for pre-ART and ART initiation date for patients on treatment in cohort years (CYs). For enrolment records, the unique identifiers allocation and/or assignment were sequential and the determination of missingness was by comparison between the list of records obtained from the electronic database and that from the pre-ART register using MS excel or MS Access. It was expected that the database lists for various cohorts were subsets of the lists in register since documentation was not real time and thus paper documentation preceded electronic data entry. In a similar manner, records of missing records of patients on treatment from the electronic records were verified against the ART register. The unique identifiers (Pepfar) and hospital numbers (facility specific) were used for the linking, verification and validation in CYs.

**The DAC process:**

Given n electronic records system facilities with $n_1$ and $n_2$ high volume and low volume patients loads respectively. Where $n_1 = n_1^1, n_1^2, \ldots n_1^k$ and $n_2 = n_2^1, n_2^2, \ldots, n_2^k$.

Given m teams ($m=m_1,\ldots,m_k$ ), then number of facilities type per team will be $n_1/m$ and $n_2/m$. Similarly, the process of access to patients records is divided into $EY_1,\ldots, EY_k$ for pre_ART and $CY_1,\ldots, CY_k$ for ART. Schematically, this process is depicted as in Figure 1.

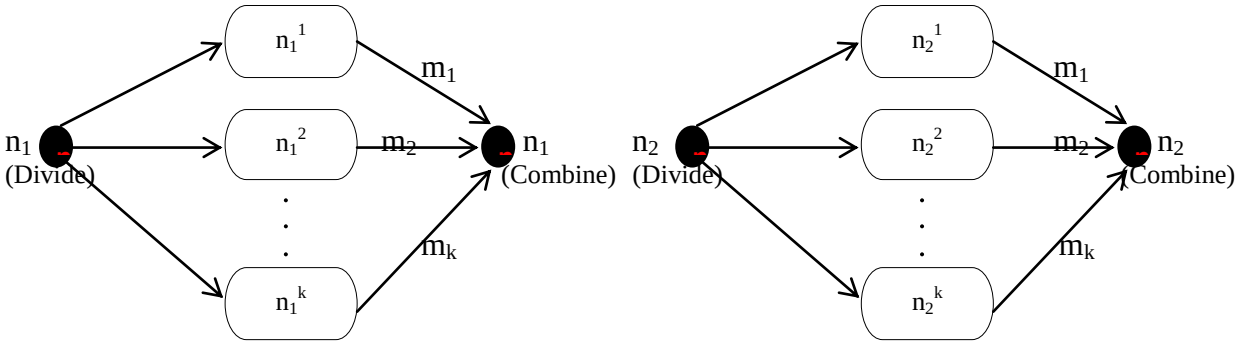


**Figure 1: Divide and Conquer approach to validation exercise**

The second stage of the validation process (see Figure 2) involved using the validated registered patients' records with identified data gaps and/or discrepancies across thematic areas-Lab orders and results, clinical encounters, pharmacy prescription and dispensing, care and support and so forth. Data gaps and other data discrepancies identified in the first and second stages were backlog to be corrected in the form of editing or data values update. Caution was exercised to avoid accumulation of more data backlog from the prospective data of current clinical encounter rendered services. A strategic information officer resident at every facility supervised data entries, extract data sets from the database and cross check data for correctness. A data management team at the central office similarly extracted data sets from different facilities databases from the central server, evaluated for completeness and other quality issues then gave feedbacks to all stakeholders in the data management system.

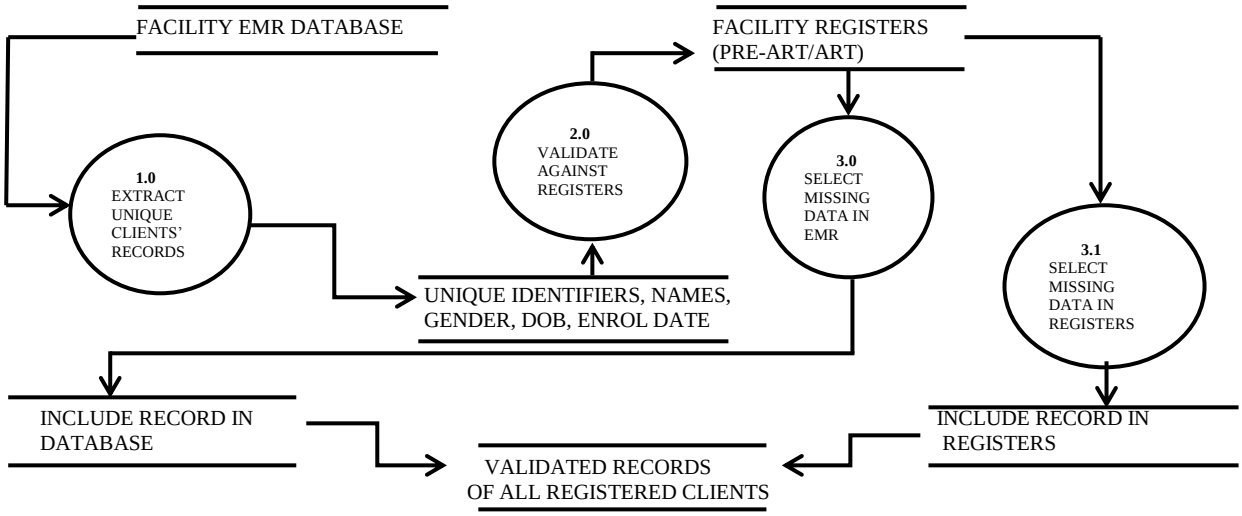


**Figure 2: On-sites validation process of registered clients**

Before initiating step 2 on a wider scale, a LQAS was applied to determine the scale of the data discrepancies. Using LQAS, sample of folders by cohorts of FYs were picked and contents were checked for data discrepancies. The aim of the LQAS approach was first hand to establish whether data discrepancies and/or anomalies existed in all or some of the facilities or none. This was to provide the team with the information to institute action for either large scale verification/validation of every folder for every facility organisation-wide or only few of the facilities. The findings were to be categorized as either "Good" or "Bad" (Figure 3). For "Good",

1. The practice was maintained at the facility and improved upon.
2. The best practice at the facility was replicated at other facilities that were below average.

On the other hand, for "Bad", the problems leading to the below average performance were identified and developed targeted solutions to improve the quality.

The following steps explain how the LQAS approach was used given that DAC was earlier used to divide all the electronic records management systems facilities into high and low volume patients (clients) load sites and divided the sites patients (clients) load into EYs:

1. Coverage: % of patients who received services from the day the facility was activated to the time of this exercise.
2. Supervision unit: The EYs
3. Supervision Area (Lot): Facility with electronic medical record system.

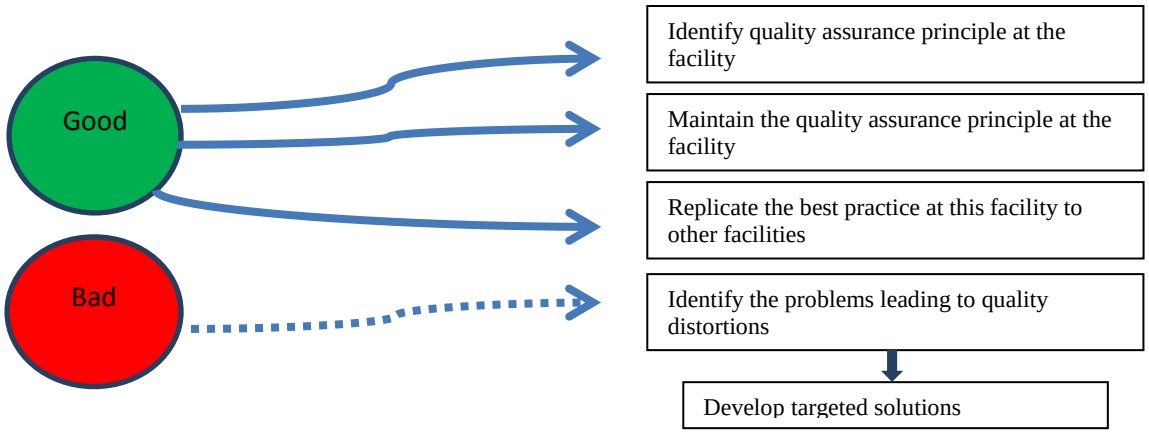


**Figure 3: Plans on using the findings**

## 2.3 Sampling Folders and Data validation

The reason for using LQAS instead of simple random sampling or any other sampling methods is because LQAS method is not too onerous, it is structured and it requires only a small sample size. The small sample size in this study is strategic because some of the contents of the folders are very voluminous especially for patients who have been in care for 10+ years and have been adherent. The use of LQAS is fit for this study since the lots are homogeneous in nature as the same services are provided with the same record filing system are used across all the facilities in the whole program. The sampling method uses probability proportionate to size (PPS), which means that sample size varies with the patients' population at facility. This include, viz:

i. Total patients' population N as illustrated in Figure 4.
ii. Using n=19, obtain number of folders across m units (EYs) of a Lot, $L_i$.
iii. Obtain sampling interval (SI) = cumulative number of patients/number in sample (N/n).
iv. Use random number table to obtain random starting number between 1 and SI.
v. Use the random starting number and sampling interval to select folders from the EYs.
vi. Divide each sampled folder content into thematic areas and chronicled them according to dates services received.
vii. Validate values of services received on filled tools against patient's records in database.
viii. Identify gaps if any and fill the gaps (see Figure 5)

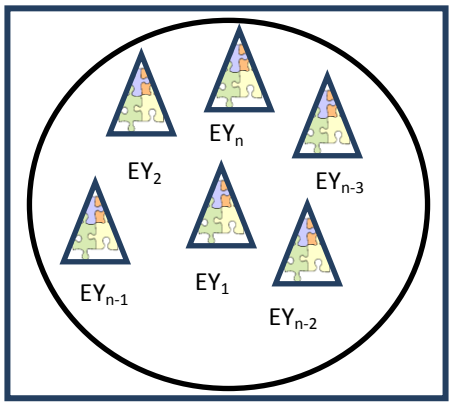


**Figure 4: Facility population view**

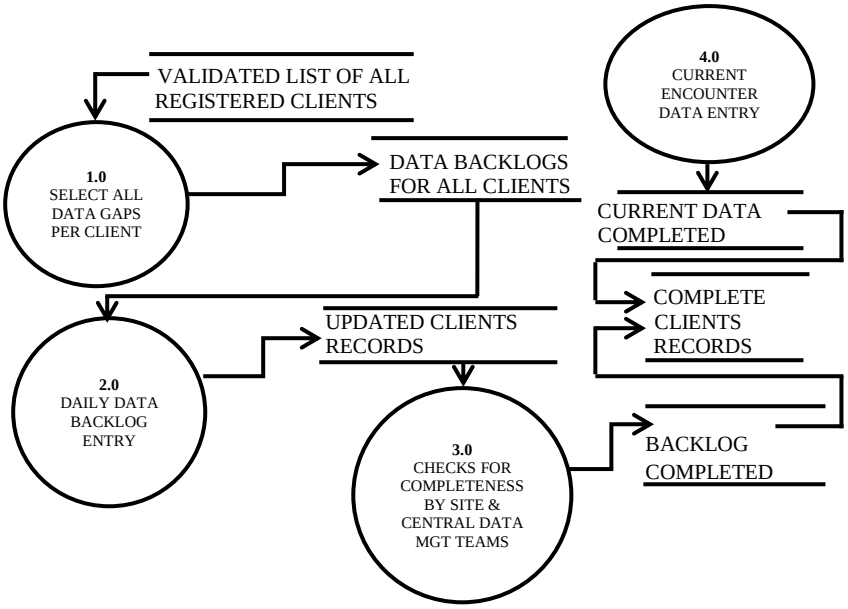


**Figure 5: Filling data gaps and entry of current encounters**

## 3. RESULTS

A revalidation exercise was carried out six months after the initial exercise to determine the level of data quality improvement. The findings from the second round of the exercise are presented in Table 1. The facilities with EMR were divided into two categories: high volume and low volume patients load.

**Table 1. Pre and Post Data Validation-Baseline and Follow-up parameters**

| | **Before validation** | | **6-month post validation** | | **Improved Quality** | |
|---|---|---|---|---|---|---|
| **Observed parameters** | **High Volume site (>=1000)** | **Low volume site (<1000)** | **High Volume site (>=1000)** | **Low volume site (<1000)** | **High Volume site (>=1000)** | **Low volume site (<1000)** |
| Number of facilities with EMR | 16 | 9 | 16 | 9 | 16 | 9 |
| % patients in registers(Pre ART and ART) not in EMR | 0.74 | 0.67 | 0.09 | 0.04 | 0.88 | 0.94 |
| % patients in EMR not in registers(Pre ART and ART) | 0.07 | 0.05 | 0.01 | 0.01 | 0.86 | 0.80 |
| % missing patient unique identifier in registers(Pre ART and ART) | 0.03 | 0.01 | 0.00 | 0.00 | 1.00 | 1.00 |
| % missing patient unique identifier (EMR) | 0.04 | 0.02 | 0.01 | 0.00 | 0.75 | 1.00 |
| % incomplete patient Identifier in registers(Pre ART and ART) | 0.01 | 0.02 | 0.00 | 0.00 | 1.00 | 1.00 |
| % incomplete patient unique identifier(EMR) | 0.05 | 0.05 | 0.00 | 0.00 | 1.00 | 1.00 |
| % missing enrolment date (Pre ART register) | 0.08 | 0.05 | 0.01 | 0.01 | 0.88 | 0.80 |
| % missing enrolment date (EMR) | 0.15 | 0.12 | 0.02 | 0.02 | 0.87 | 0.83 |
| % missing DOB in registers(Pre ART and ART) | 0.40 | 0.34 | 0.04 | 0.01 | 0.90 | 0.97 |
| % missing DOB (EMR) | 0.44 | 0.42 | 0.05 | 0.06 | 0.89 | 0.86 |
| % missing ART initiation date (ART register) | 0.06 | 0.08 | 0.02 | 0.01 | 0.67 | 0.88 |
| % missing ART Initiation date (EMR) | 0.25 | 0.17 | 0.03 | 0.01 | 0.88 | 0.94 |
| % missing follow-up visits updated in registers(Pre ART and ART) | 0.33 | 30.5 | 0.04 | 0.04 | 0.88 | 0.89 |
| % missing follow-up visits updated (EMR) | 0.45 | 0.32 | 0.10 | 0.08 | 0.78 | 0.75 |

| | | | | | | |
|---|---|---|---|---|---|---|
| % wrong follow-up visits updated in registers(Pre ART and ART) | 0.02 | 0.02 | 0.00 | 0.00 | 1.00 | 1.00 |
| % wrong follow-up visits updated(EMR) | 0.08 | 0.10 | 0.02 | 0.01 | 0.75 | 0.90 |
| % missing patients names in registers (Pre ART and ART) | 0.04 | 0.06 | 0.01 | 0.01 | 0.75 | 0.83 |
| % missing patients names (EMR) | 0.05 | 0.03 | 0.02 | 0.01 | 0.60 | 0.67 |
| % missing sex in registers(Pre ART and ART) | 0.02 | 0.01 | 0.00 | 0.00 | 1.00 | 1.00 |
| % missing sex in registers(EMR) | 0.06 | 0.04 | 0.01 | 0.01 | 0.83 | 0.75 |
| % missing baseline CD4 count (Pre ART and ART) | 0.10 | 0.13 | 0.01 | 0.01 | 0.90 | 0.92 |
| % missing baseline CD4 count (EMR) | 0.14 | 0.12 | 0.02 | 0.03 | 0.86 | 0.75 |
| % missing follow-up CD4 count (Pre ART and ART) | 0.35 | 0.30 | 0.04 | 0.03 | 0.89 | 0.90 |
| % missing follow-up CD4 count (EMR) | 0.40 | 0.42 | 0.05 | 0.02 | 0.88 | 0.95 |
| % missing WHO clinical stage baseline in registers(Pre ART and ART) | 0.15 | 0.11 | 0.01 | 0.01 | 0.93 | 0.91 |
| % missing WHO clinical stage baseline (EMR) | 0.17 | 0.10 | 0.02 | 0.02 | 0.88 | 0.80 |
| % missing WHO clinical stage follow-up(Pre ART and ART) | 0.16 | 0.13 | 0.03 | 0.01 | 0.81 | 0.92 |
| % missing WHO clinical stage follow-up(EMR) | 0.18 | 0.15 | 0.04 | 0.04 | 0.78 | 0.73 |
| % duplicate records (EMR) | 0.15 | 0.08 | 0.02 | 0.01 | 0.87 | 0.88 |
| % untraceable patients folders | 0.04 | 0.02 | 0.02 | 0.01 | 0.50 | 0.50 |

## 4. DISCUSSIONS

The results obviously indicated a tremendous loss of patients' records in the EMR with an average of 70.5% as compared to the paltry absence of an average of 6% from both hardcopies registers. The current sites EMR databases or repositories were derived from the databases of the two previously implemented EMRs. The discrepancies, majorly, were due to issues identified with data migration technicalities, irregularities associated with the second EMR- Teleform, implemented prior to Open MRS and also due to inconsistency of policy or procedures on data entry. Data migration is being practiced in data management sphere, for instance, the consultative committee for space data systems recognized and used amongst its other objectives the migration of digital information to new media and forms and the role of software in information preservation (CCSDS, 2012). For the 6% on average of records found in EMR but missing in the registers, it was observed that this occurred as a result of omissions. The data capturing tools are forms which are used to enter data into the electronic platform and the registers. Data entry into the electronic records systems is not real time. The forms used to enter data into the electronic platform were not all passed on to be used for entries into the registers.

The bulk of the discrepancies for almost all parameters were centered on the EMR (see Figure 6). The average discrepancies witnessed with the EMR before validation ranged from 3-70.5% with median (IQR) of 13.5%(34%) as compared to that of the registers 0-37% with median (IQR) of 14.5%(13%). Similarly, a look at the post validation indicated that the average discrepancy in EMR was still higher although substantially reduced to a very low percent range of 0-6.5%. On the other hand the average discrepancies recorded in the registers was lower with a range of 0-4%.

A statistical comparison test of the pre and post validation outcomes showed a tremendous data quality improvement. As shown in Tables 2 and 3 and Figure 6, the mean (CI, 95%) differences for EMR and registers were 17.67 [6.94599, 28.38734] and 10.37 [3.340688, 17.39265] respectively. There was a significant difference in the impact of the method used to improve the data quality for both the EMR and registers with $p = 0.0032$ and $p=0.0068$ in that order. From table 1, a huge success in identifying and initiating the process of data quality was recorded as least quality gained was 50% and the greatest 100%. The lowest was attributable to lost folders that were mostly in the facilities with more number of elites. In addition, the impact of the quality improvement method was higher at the low volume facilities (expected) as compared the corresponding high volume facilities. The average(mean), average (median) and IQR ratios of high volume to low volume quality improvement were 1:1.02 (0.86/0.88), 1:1.02 (0.88/0.90) and 1:1.89(0.09/0.17) in that order.

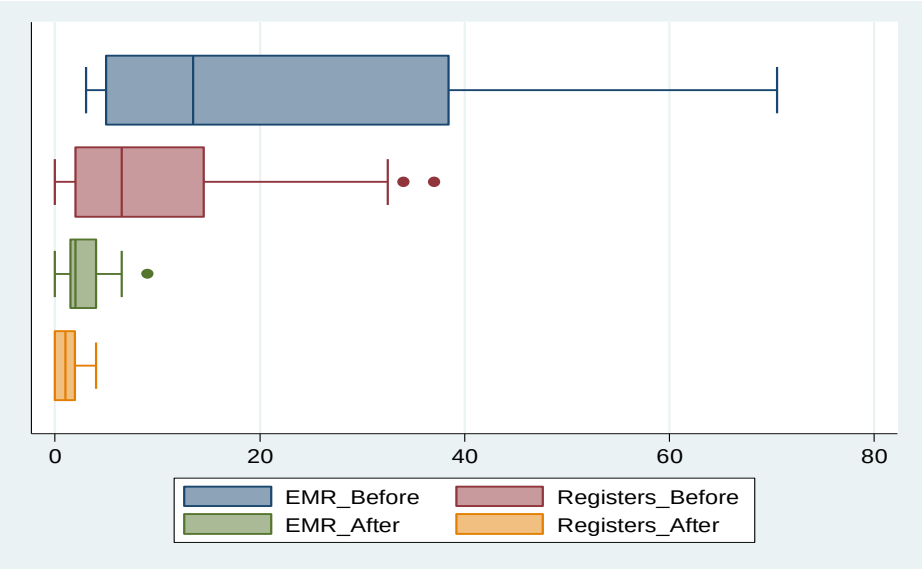


**Figure 6: Average percent discrepancies**

**Table 2. EMR pre and post validation of data**

```
. ttest EMR_Before == EMR_After, unpaired unequal

Two-sample t test with unequal variances

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
EMR_Be~e |      15    20.53333    4.972895    19.25994    9.867535    31.19913
EMR_Af~r |      15    2.866667    .6352103    2.460159    1.504276    4.229057
---------+--------------------------------------------------------------------
combined |      30        11.7    2.959264    16.20855    5.647626    17.75237
---------+--------------------------------------------------------------------
    diff |            17.66667      5.0133                6.94599    28.38734
------------------------------------------------------------------------------
    diff = mean(EMR_Before) - mean(EMR_After)                     t =   3.5240
Ho: diff = 0                     Satterthwaite's degrees of freedom =  14.4567

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9984         Pr(|T| > |t|) = 0.0032          Pr(T > t) = 0.0016
```

**Table 3. Hard copies -registers pre and post validation of data**

```
. ttest Registers_Before == Registers_After, unpaired unequal

Two-sample t test with unequal variances
```

| Variable | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Regist~e | 15 | 11.6 | 3.26533 | 12.64657 | 4.596563 | 18.60344 |
| Regist~r | 15 | 1.233333 | .3304638 | 1.279881 | .5245589 | 1.942108 |
| combined | 30 | 6.416667 | 1.877894 | 10.28565 | 2.575943 | 10.25739 |
| diff | | 10.36667 | 3.28201 | | 3.340688 | 17.39265 |

```
    diff = mean(Registers_Before) - mean(Registers_After)         t =   3.1586
Ho: diff = 0                     Satterthwaite's degrees of freedom =  14.2868

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9966         Pr(|T| > |t|) = 0.0068          Pr(T > t) = 0.0034
```

## 5. CONCLUSION

Drawing from the results, it was observed that the magnitude of the data discrepancies came from the electronic medical records system which invariably was not unconnected with the non- sticking to one patient record management system and policy. If the problems were from the data capture and summary tools, the discrepancies could have been attributed mainly to other factors such as the stakeholders/personnel involved, lack of data management procedures, and so forth. By this, it is safe to say that maintaining a stable robust electronic patient's records management system will reduce data errors and loss of patient's records and/or vital information. This finding may not be limited only to health records systems but to all data management systems. One of the factors noted (data from current EMR –Open MRS compared with first EMR-careware) was the data migration technicality flaws.

In as much as there is need to maintain hard copies of data, it is wise to state that real data entry at the service delivery points by the service provider may reduce loss of data through omission as was seen in the gaps for data found in EMR that were missing in hard copies – registers. These little omissions gradually in the long run will amount to big loss of data (Gibney and Van Noorden, 2013). Even though, the authors refer to loss of improperly archive data but it is applicable to improperly collected and collated data. Vines et al., (2014) on availability of research data also expressed concerns on data that are being lost at alarming rate. Not that the data were not captured on the data capturing tools (forms) but they were not entered into the registers which are the summary tools for all patients that were provided with one service or the other. Transitioning from one data management system to the other, if not carefully studied, framework well modeled, and implementation plans diligently executed, may always make mess of the new system. It is therefore safe to say that policy changes regarding to EMR may be attractive especially with changing technologies and increase in data along with organizational size but it may be inessential when data quality is compromised or data loss is incurred.

## 6. ACKNOWLEDGEMENTS

We thank the management of Institute of Human virology Nigeria for giving us the opportunity to carry out this work. We also appreciate the efforts and assistance of Nifarta Andrew, Eunice Ekong and Bidemi Harry_Erin and Fati Murtala Ibrahim, all of the Department of Strategic Information department, Institute of Human Virology Nigeria.

## 7. REFERENCES

[1] Ajami, S. and Bagheri-Tadi, T. 2013. Barriers for Adopting Electronic Health Records (EHRs) by Physicians. Acta Inform Med. 2013; 21(2): 129–134.Published online 2013 Jun. doi: 10.5455/aim.2013.21.129-134

[2] Blelloch, G. 2011. Parallel and Sequential Data Structures and Algorithms. Lecture 15-210 (Fall 2011).

[3] CAREWare. Retrieved from http://www.jprog.com/wiki/All-CAREWare-documentation.ashx).

[4] Consultative Committee for Space Data Systems. 2012. *Reference model for an Open Archival Information System (OAIS)* (Magenta Book CCSDS 650.0-M-2). Retrieved from http://public.ccsds.org/publications/archive/650x0m2.pdf

[5] Elamparithi, M. and Anuratha, V. 2015. World Journal of Computer Application and Technology 3(3): 41-48. Retrieved from http://www.hrpub.org DOI: 10.13189/wjcat.2015.030301.

[6] Gee, M. and Helfert, M. 2013. Cost and Value Management for Data Quality. Handbook of data Quality pp 75-92.

[7] Gibney, E. and Van Noorden, R. 2013. Scientists losing data at a rapid rate. *Nature*. doi:10.1038/nature.2013.14416. Retrieved from http://dx.doi.org/10.1038/nature.2013.14416.

[8] Herring, M. 2016. Data Is Your Most Valuable Asset. Retrieved from https://dzone.com/articles/data-is-your-most-valuable-asset.

[9] Horthonworks 2016. Data Is Your Most Valuable Asset. Retrieved from http://hortonworks.com/blog/data-is-your-most-valuable-asset.

[10] Hund, L.2014. New tools for evaluating LQAS survey designs. DOI**:** 10.1186/1742-7622-11-2.

[11] IBM 2016. Data management: Tapping your most valuable asset. Retrieved from https://www.ibm.com/midmarket/us/en/article_DataManagement2_1209.html.

[12] Lin, B. and Chan H.G. 2000. Managing data quality in the health care industry: Some critical issues. Journal of International Information Management, Vol. 9 [2000], Number 1. CSUSB Scholar Works.

[13] MIT Technology Review Custom 2016. Produced in partnership with Oracle. Retrieved from www.technologyreview.com/media.

[14] Peer, L., Green, A, and Stephenson E. 2014. Committing to Data Quality Review. International Journal of Digital Curation 2014, Vol. 9, Iss. 1, 263–291.

[15] Piette, J.D, Lun, K.C, Moura, L.A., Fraser, H.S.F., Mechael, P.N., Powell, J., & Khoja, S.R. 2012. Impacts of e-health on the outcomes of care in low- and middle-income countries: where do we go from here? *Bulletin of the World Health Organization* 2012;90:365-372. doi: 10.2471/BLT.11.099069.

[16] Shankaranarayanan, G. and Blake, R. 2017. From Content to Context: The Evolution and Growth of Data Quality Research. Journal of Data and Information Quality (JDIQ). Vol.8 Issue 2. ACM New York. Retrieved from http://dl.acm.org/citation.cfm?doid=3035914.2996198.

[17] Sadiq, S., Khodabandehloo,N., Indulska,Y.M. 2010. 20 Years of Data Quality Research: Themes, Trends and Synergies.

[18] Talc. 2003. Assessing Community Health Programs: Using LQAS for Baseline Surveys and Regular Monitoring.

[19] Teleform. Retrieved from https://en.wikipedia.org/wiki/TeleForm.

[20] Teradata 2013. Breaking the Application Barrier: Why Data is the Most Valuable Asset in the Oil and Gas Industry. Retrieved from http://assets.teradata.com/resourceCenter/downloads/WhitePapers/EB6802.pdf

[21] USAID and NUMAT 2010. LQAS Survey Report: A Household Survey on Malaria, HIV&AIDS and TB Interventions in Nine Districts of Northern Uganda.

[22] Vines, T.H., Albert, A. Y.K., Andrew, R.L., Débarre, F., Bock, D.G., Franklin, M.T., Rennison, D.J. 2014. The availability of research data declines rapidly with article age. *Current Biology 24*(1), 94-97. doi:10.1016/j.cub.2013.11.014.

[23] Wang, R.Y. and Strong, D.M. 1996. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems 12*(4), 5-33. Retrieved from http://mitiq.mit.edu/Documents/Publications/TDQMpub/14_Beyond_Accuracy.pdf

[24] Warme, T. 2014. Data: your most valuable asset. A business case for data governance. Retrieved from https://www.linkedin.com/pulse/20141119183044-243338813-data-your-most-valuable-asset-a-business-case-for-data-governance.