Clinical Medicine Research
2016; 5(4): 63-72
http://www.sciencepublishinggroup.com/j/cmr
doi: 10.11648/j.cmr.20160504.11
ISSN: 2326-9049 (Print); ISSN: 2326-9057 (Online)

A Test of Accuracy and Completeness in Data Flow Value Chain Within the Context of Early Infant Diagnosis of HIV Using Mixed Methods Approaches

Leonard King'wara[1, *], David Omondi Okeyo[2], Charles Obonyo[3]

[1]Department of Public Health, Maseno University, Kisumu, Kenya

[2]Schools of Public Health, Maseno University, Kisumu, Kenya

[3]Kenya Medical Research Institute, Kisumu, Kenya

Email address:

leonard.kingwara@gmail.com (L. King'wara), jandigwa@yahoo.co.uk (D. O. Okeyo), cobonyo@yahoo.com (C. Obonyo)

[*]Corresponding author

To cite this article:

Leonard King'wara, David Omondi Okeyo, Charles Obonyo. A Test of Accuracy and Completeness in Data Flow Value Chain Within the Context of Early Infant Diagnosis of HIV Using Mixed Methods Approaches. *Clinical Medicine Research*. Vol. 5, No. 4, 2016, pp. 63-72. doi: 10.11648/j.cmr.20160504.11

Received: May 5, 2016; **Accepted**: May 17, 2016; **Published**: June 4, 2016

Abstract: Recent changes to Kenya's early infant diagnosis (EID) testing algorithm have raised hope that the national goal of reducing perinatal HIV transmission rates to less than 5% can be attained. While programmatic efforts to reach this target are underway, obtaining complete and accurate data from clinical sites to track progress presents a major challenge. The purpose of this study was therefore to assess data completeness, accuracy and challenges in relation to routine EID data management in Kisumu East and Kisumu West Districts within Kisumu County, Kenya. Purposive sampling was used to select 23 active health facilities across the two districts disproportionate by partners. From the selected facilities a sampling frame of 192 patients was established and a sample size of 130 patient's data selected. Accuracy and completeness were determined by computing sensitivity, specificity, and predictive values. Infant sex, Infant prophylaxis, Breastfeeding information, Entry point, and Test results were found to have registered replication completeness not significantly different from 99% replication at ($p>0.05$) as oppose to Sample code, Infant Age, Date of sample collection, and PMTCT prophylaxis were found to be significantly different from the 99% replication ($p<0.05$). On completeness; Sample code and date of sample collection registered completeness beyond the hypothesized value, implying they were complete as required. The remaining data elements including infant sex, infant age, infant prophylaxis, PMTCT prophylaxis, breastfeeding information and entry point were significantly lower than the hypothesized completeness value ($p<0.0001$) except for test results that had completeness score equal to 0.99. The study, therefore, concludes that despite the shortage of staffing and other challenges, personnel working in the data management system appear to be dedicated, informed and conscientious. However, this research suggests that there is a suboptimal use of the information for local action in certain areas. This assessment thus serves to enlighten policy-makers on the current state of the EID data management system in Kisumu East and West districts.

Keywords: Data Accuracy, Completeness, Early Infant Diagnosis, HIV, PMTCT

1. Introduction

There are indications that healthcare service delivery is still weak due to lack of accurate and credible data to inform essential health products supply and this has continued to cost lives [1] especially among children at risk of HIV injuries. Key causes to these are gaps and weaknesses in demand forecasting resulting in a mismatch of data and information [2]. In Kenya, these have resulted in children missing out on malaria medicines, pregnant mothers not sleeping under impregnated bed nets, and those living with HIV missing their medicine cycles, thus jeopardizing their lives and contributing to the risk of viral drug resistance within their community, among others [3]. All this point to a deficiency in health information systems, that should be used

in informing managerial decision making and improving operations in areas such as epidemiologic surveillance, health outcome assessment, program and clinic administration, program evaluation, performance measurement, public health planning and policy analysis.

According to current statistics over 67% of Kenyans deaths are due to health information systems failures in health care [3]. The Early Infant Diagnosis (EID) program, which is one of the several health programs aimed at monitoring the impact of PMTCT (Prevention of Mother to Child Transmission) is among the health information systems characterized by incoherent and incomplete data [4]. This has forced most decision makers to use approximation and at times guesswork leading to inadequate allocation, over allocation or wastage of resources that would have otherwise been channeled towards other important livelihood factors [3]. As a result of inaccurate health records, the Kenyan government invested heavily in the pharmaceuticals but failed to invest in non-pharmaceuticals in 2006 due to lack of funds thus putting patient care in jeopardy [3]. This buttress the fact that about 75% of patients and health care providers considered the data records within the health facilities to be fragmented, convoluted, and plagued by duplication of effort, poor communication and conflicting advice [2]. These inefficiencies have resulted in double digit inflation of cost of health care worldwide thus making it almost impossible for an individual or even governments to provide health services adequately to its citizenry [5]

World Health Organization (WHO) had a long time ago indicated that health information systems were criteria to reach "health for all" in 2020 which also emphasized that advanced health information systems were required for an effective health sector management [6]. In this context, most Ministry of Health all over the world made a decision to use the information systems as a tool for increasing effective management of the health services. These systems can either be computerized or manual, and regardless of nature or type, it must meet certain standards as defined by various countries. An optimum health information system management is intended to increase the efficiency of health services using fewer resources with the primary goal being minimization of losses in services at the effectiveness level [3]

EID health information system in Kenya is relatively new and was initiated during the rapid scale-up of EID program and the driver for its initiation was the need for strong systems of centralized data at the Ministry of Health [7]. This could enable assessment of the national program by monitoring of progress and inform interventions which were required in strengthening the program [8]. Program assessments in turn revealed the need for interventions at the facility level on how to improve care and retention outcomes, including improved data tools for linking & tracking HIV-exposed infants [7]. These new facility-level data tools were then integrated with central data capture tools at MOH to improve linkage of the patient to care, track patients at the facility level, develop stronger systems for data flow between facilities and MOH, and increase capacity for robust program analysis at the MOH [4].

Despite these good intentions, three important factors have remained as key challenges to EID health information system in Kenya. First, the integration of key EID indicators into PMTCT monthly reporting form that facilities should complete remains a key challenge**.** Second, feedback mechanism from the Ministry of Health back to health centers also remains a major a concern. Finally, the integration of EID services into district health management to enable the involvement of the District Health Management Team (DHMT) in the EID program also remain a challenge [9]. All these factors appear to have an influence on the accuracy and consistency of the data generated by EID health information system.

Due to these challenges associated with the EID programme health information system, we suspected inconsistencies and inaccuracy of data management system supporting this information system. It is on this basis that the study evaluated routine health system data management in EID programme within Kisumu East and West districts. The purpose of evaluating EID programme was therefore to ensure that problems under surveillance (PMTCT success) are monitored efficiently and effectively. The programme should then be evaluated periodically to determine how well they operate to meet their stated purposes and objectives. Evaluation findings yielded recommendations for improving surveillance quality, efficiency, and usefulness. This article therefore explores completeness, accuracy and challenges associated with EID data management in Kisumu East and West district within Kisumu County in Kenya.

2. Materials and Methods

2.1. Study Design

The study adopted a cross-sectional survey where data was collected once and analysed. Specifically, the survey adopted concurrent nested mixed methods approaches where both quantitative and qualitative data were collected and analysed with quantitative data being given priority and qualitative data used to triangulate the outcome of quantitative data. The approach allowed the study to simultaneously generalise results from sample to population and gain a deeper understanding of the EID health information system.

2.2. Population and Sampling of Patient Records

The population of the study targeted data records for 192 mothers and children attending PMTCT services in 23 health facilities within Kisumu East and West district within Kisumu County. Eligibility criteria included a child who has attended the respective facility for EID services between the month of May and August 2012, below 18 months of age and a health care worker dealing with EID services between the periods of the study. The sample size was determined using Yamane which is a simplified formula for proportions [13]. The sample size is denoted by n while the population is N (sampling frame). $n = N/1+N\ (e)^2 = 192/1+192\ (0.05)^2$. A

population of 192 patients data thus gives 130 patients data [14]. Sampling procedure involved the use of simple random technique where all the eligible infants were given equal chance of being involved in the sample. This was done based on unique code numbers for mother and child. For qualitative phase theoretical (purposive) sampling technique was used to select health care workers to saturation. The selection was based on good knowledge of EID programmes and operations.

2.3. Data Collection and Analysis Plan

Quantitative data collection: The questionnaire was structured to include all the relevant EID data elements under focus in the study. These data elements included the number of tested infants, tested infant age, maternal PMTCT regimen, 1st and 2nd PCR results, date of sample collection, reception at the central testing laboratory and dispatch from the testing laboratory. *Qualitative data collection*: An interview schedule was used to explore the challenges associated with EID data management processes. The questionnaire included all the relevant information in the EID data management system and programmatic activities influencing data management. All health care workers selected were interviewed and questioned about aspects of the EID data management system description.

The initial preparation of EID data management system information for statistical analysis involved the pairing of raw data from the facilities with the corresponding data from the central testing laboratory in an excel spread sheet. The pairing was to help review the data for accuracy and completeness. Data analysis involved both the descriptive and the inferential statistics. Descriptive statistics was meant to provide background information of the infants under the study. This was done using SPSS version 19.0

To analyse data for completeness, each data was retrieved in separate excel sheets. The hypothesised completeness per each data element was then set at 99%. Completeness per each data element was then determined by getting the proportion of all the complete data follows:

$$\text{Proportion complete (p)} = \frac{\text{Number of complete data elelemts}}{\text{Total number of data elements of interest}}$$

Proportions for complete data per each data element was then determined for both the soft and central testing laboratory data. Z score statistics was then performed to determine how significantly different the proportion achieved was from the hypothesised value of 99% at 95% confidence level.

To analyse data accuracy the level of accuracy was set at 99%. For accuracy only 1% was allowed for inaccuracy. Accuracy was looked at in terms of replicability of the raw data in the soft data collected from the central testing laboratory. The proportion of replication was determined for each of the data elements as follows:

$$\text{Proportion accurate (p)} = \frac{\text{Total number of data replicated in both raw and soft data in the testing lab}}{\text{Total number of data elelments being correlated}}$$

A two z score was then done using SPSS Version 19.0 to establish whether the proportion generated and the 99% targeted were significantly different at 95% level of confidence.

Qualitative data analysis adopted the styles of Framework analysis for the KIIs information. One key advantage with this framework analysis is that although it uses a thematic approach, it allows themes to develop both from the research questions and from the narratives of research participants. The process of data analysis began during the data collection, by skillfully facilitating the discussion and generating rich data from the interviews, complementing them with the observational notes and typing the recorded information. This stage was be followed by familiarization with the data, which was achieved by reading transcripts in their entirety several times and reading observational notes taken during the interview. The aim was to immerse in the details and get a sense of the interview as a whole before breaking it into parts. During this process major themes began to emerge. The next stage involved identifying a thematic framework, by writing *memos* in the margin of the text in the form of short phrases, ideas or concepts arising from the texts and beginning to develop categories. At this stage descriptive statements were formed and an analysis carried out on the data under the questioning route. The third stage, *indexing*, comprised sifting the data, highlighting and sorting out quotes and making comparisons both within and between cases. The fourth stage, charting, involved lifting the quotes from their original context and re-arranging them under the newly-developed appropriate thematic content [15].

Approval of the research proposal was given by the Maseno University Research ethics committee. Ethical principles such as anonymity were assured in the data by use of codes that could not actually identify participants by name. Data related to the participants were stored in excel soft copy in a computer with a password so that nobody could access it.

3. Results

3.1. Descriptive Statistics of Early Infant Diagnosis Patient Data

The data was derived from 13 health facilities in Kisumu East and West districts. It had patient data of Infants attending PMTCT whose age range from 2 months to 18 months. The median age of the infant was 6 months while the

mean age was 7.5 months. The majority of the infants were exclusively breastfed for six months. The PMTCT regimen of choice comprised of AZT +NVP+3TC, sdNVP, and HART while the common entry point for these infants into the facility was MCH clinic. Patient data collected consisted of 60 male infants, 65 female infants and 5 unidentified genders in the patient registrar. Translating to 50% female, 46.1% male and 3.9% unidentified sex for infants.

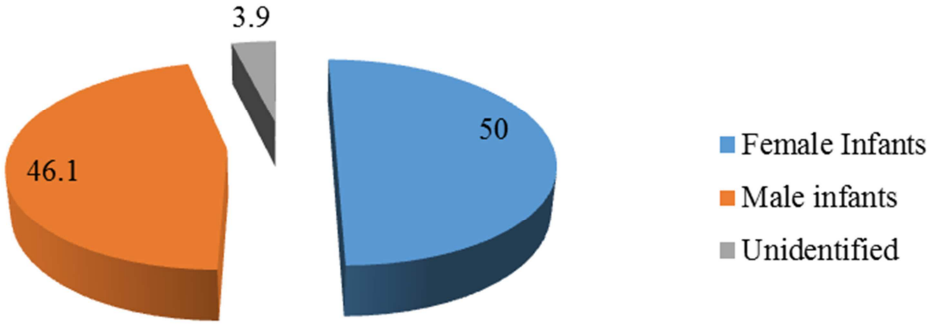

Figure 1. *Distribution of the infants by sex.*

3.2. Accuracy of Data in Kisumu East and West District

The accuracy of data was based on replication between the facility raw data and the central testing Lab soft data. Using a two tailed hypothesis on a test of one proportion at a significance level of α=0.05, the researcher determined the significance of the difference in replication for between the facility and the central laboratory data as indicated below.

Table 1. *Single proportion z-score test for the level of accuracy of the data element.*

Data element (n=130)	Proportion replicated	Z value	P value	95% CI
Sample code	0.97	2.9382	0.0171	0.92-0.99
Infant Sex	0.98	0.619	0.5361	0.95-1.00
Infant Age	0.94	5.901	< 0.0001	0.88-0.97
Infant prophylaxis	0.98	1.501	0.1333	0.93-0.99
Date of sample collection	0.97	2.384	0.0171	0.92-0.99
PMTCT prophylaxis	0.96	3.266	0.0011	0.91-0.98
Breast feeding information	0.98	1.501	0.1333	0.93-0.99
Entry point	0.98	1.501	0.1333	0.93-0.99
Test results	0.98	1.501	0.1333	0.93-0.99

Table 1 estimated how far the proportions of replicated data were from the hypothesized value of 99% for all the data elements under investigation. Infant sex, Infant prophylaxis, Breast feeding information, Entry point, and Test results were found to have registered proportion replicated completeness not significantly different from 99% replication at ($p<0.05$). The proportion replicated for the remaining data elements like Sample code, Infant Age, Date of sample collection, and PMTCT prophylaxis were found to be significantly different from the 99% replication ($p<0.05$). This implies that out of the nine data elements, four can be classified as inaccurately recorded.

3.3. Completeness of Two Data Sets

Figure 2 below shows how far by percentage the data elements are far from the ideal 100% complete. Sample code and date of sample collection are the most completely filled data elements recording 100% completeness. Based on the targeted completeness of 99% sample code, and date of sample collection were the only data elements that achieved the 99% mark. The remaining data elements recorded below the 99% depicting some level of incompleteness.

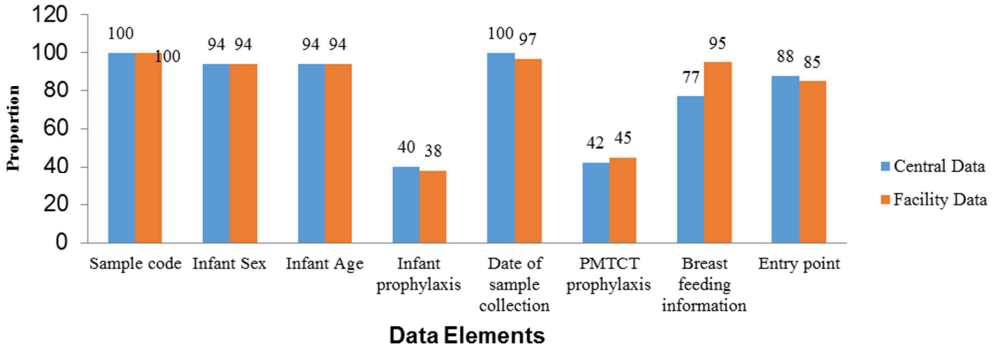

Figure 2. *Distribution of data elements by proportion of completeness.*

However, to determine how significant incompleteness was a two tailed hypothesis test of single proportion at a significance level of α=0.0 was used for both the facility raw data and the central laboratory soft data as indicated below.

Table 2. *Single proportion z-score test for the level of completeness of data element in the facility laboratory N=130.*

	Facility data				Central Lab data			
Data element	**Proportion of complete data**	**Z value**	**P value**	**95% CI**	**Proportion of complete data**	**Z value**	**P value**	**95% CI**
Sample code	>0.99	n/a	n/a	n/a	>0.99	n/a	n/a	n/a
Infant Sex	0.94	5.730	****	0.89-0.97	0.94	5.73	****	0.89-0.97
Infant Age	0.94	5.730	****	0.88-0.97	0.94	5.73	****	0.89-0.97
Infant prophylaxis	0.38	69.901	****	0.30-0.47	0.40	67.609	****	0.32-0.49
Date of sample collection	0.97	2.292	*	0.93- 0.99	>0.99	n/a	n/a	0.97-1.00
PMTCT prophylaxis	0.45	61.880	****	0.37-0.54	0.42	65.317	****	0.34-0.51
Breast feeding information	0.75	27.502	****	0.67-0.82	0.77	25.210	****	0.69-0.83
Entry point	0.85	16.043	****	0.78-0.90	0.88	12.605	****	0.81-0.92
Test results	0.96	3.438	***	0.91-0.98	0.99	0.000	P=1.0000	0.95-1.00

*****=p<0.0001; ***=0.05>p>0.0001; *=p>0.05*

Table 2 estimated how far the proportions of complete data were from the hypothesized value of 99% for all the data elements under investigation. Sample code and date of sample collection registered completeness beyond the hypothesized value, implying they were complete as required. The remaining data elements including infant sex, infant age, infant prophylaxis, PMTCT prophylaxis, breast feeding information and entry point were significantly lower than the hypothesized completeness value ($p<0.0001$) except for test results that had completeness score equal to 0.99. This implies that the entry of data at both the central lab and the facility raw data were incomplete as would have been expected except for the sample code, date of sample collection and test results.

3.4. Challenges Associated with Data Management System Attributes

During analysis of the qualitative data, a number of thematic issues emerged as the key challenges to EID data management system. Among the themes are data quality, the level of usefulness, the simplicity of the system, flexibility, acceptability, positive predictive value, representativeness and stability. The above themes outlined emerged as key areas that would explain the challenges associated with health data accuracy and completeness.

From a voice perspective *quality of data* as a theme reflected the completeness and validity of the data recorded in the EID data management system. The evaluation process established a strong voice on poor quality data collection to be related to insufficient staffing. Double work done by the clinical staffs in handling both the patients and data work was a serious threat to data quality. Sometimes they would get assistance from the supporting partner who would at times employ a data clerk, but this assistance is withdrawn at times based on the priority of the partners thus compounding the staffing challenges as a serious threat to data quality. Apart from staffing, timely reporting also emerged as some of the voices that threaten the quality of data collection. Delays in data reporting characterised by a complete lack of reporting in certain months interfere with the quality of data being collected. This has even necessitated following up by the PMTCT coordinators to follow ups to check on the data quality as evidenced in the quote below. A critical quote reflects the respondents' opinion on the absence of a dedicated data-capturer, the reporting frequency timeliness and completeness of data submission was seldom checked. Respondents further stated that report submissions from facilities were erratic but that specific follow-up into reporting practices was done on an ad hoc basis on suspicion of irregularities.

> 'We don't have a dedicated data capture……. Most of the time the partner employs someone to capture the data. It is impossible for me to handle the patients and fill in all those data elements. If the partner decides and withdraws their person then it becomes difficult to fill all those data elements alone said the MCH in charge'

> 'Most of the facilities do not report all the data in a timely manner. In some cases, they report for certain months and leave out others. We try to do follow up in specific facilities in case we are suspecting irregularities said the PMTCT coordinator'

The level of usefulness of a system is determined by how the system fulfils its intended objectives as to why it was established. For a system to fulfil its objectives, the said objectives should be clear to both the developer of the system, the system administrators, and the users of the system and finally the consumers of the data generated by the system. During analysis, Lack of understanding of the objectives of the system by its users emerged a threat to the usefulness of the EID data management system. This was voiced in the quote where all three key respondents indicated that surveillance data have never been used to change policies and procedures. Respondents even wondered whether that data is being analysed and what is being done by the data thereafter as indicated by their question of one of the respondent

> 'Why do you need all these while you can just use the total number of patients to do procurements asked the facility lab in charge?'

> 'Which policies are these that are supposed to be changed

by the data collected? Inquired one of the PMTCT coordinator'

Most thought that the data was being used in procurement of the requirement for EID test kits and consumables alone. Additional users of the data identified were the Management science of health and the World Health Organisation. The two organisations use the system for procurement of consumables.

'We also submit these data to Management science of Health and NASCOP which shares it with WHO said one of the PMTCT coordinators'

The easiness system usage and of various components of a system communicating continuously with each other without breakage defines the system *simplicity*. At analysis, double standards in the mode of communication emerged as a serious threat to the EID data management system simplicity. The processes of EID data management system starts with the collection of all the data elements by a health practitioner for all HIV exposed infants. Data so generated is passed through a series of statutorily defined routes through the different levels of health care to the NASCOP. This double standard in the information flow pathway is indicated by the quote below that tries to define different modes of information flow depending as to whether it is a private or a public facility. The patient's initial point of care is the site from which reporting takes place. The routes of notification vary between public and private sectors. The respondents all indicated that MTCT case reporting from facility to district level and from the district level to provincial level occurred by report collection by the PMTCT coordinator. Respondents stated that some facilities report cases directly to the provincial offices while others send the PMTCT reporting forms through to the DASCO. The DASCO then submit notification data to the provincial coordinator as indicated by the response. With such inconsistent information flow path, any change of staff or in a case where the responsible staff goes on leave will lead to breakage in the information as evident in the quote below.

'We submit the data directly to the provincial office, but some facility in the lower levels submit to the DASCO who then submit the data to the provincial coordinator-said the nurse in charge'

'Private facilities do not normally submit data. It is the duty of the PMTCT coordinator and the partner in the area to go and obtain data personally from them and this at times leads to delay in data submission when there is a new staff or the responsible staff going on leave said one of the PMTCT coordinators'

From time to time based on the changing needs of the programme or surveillance, a need for changing certain component of the system that may be considered redundant or adding some that are considered necessary may occur. This ability to change based on the emerging needs defines the *flexibility* of the system. During analysis, lack of funds and dependency on donor emerged as clear voices in the impediment to EID data management system flexibility. As evident in the quote below, most respondents indicated that changing the official national MTCT list of data elements is cumbersome and since the process is paper-based and not electronic at the lower levels of health facility presently, any alterations require reissuing of MTCT data elements guidelines to all health care providers. This is most likely to be an uphill task as the EID data management system is heavily donor dependent and any alteration of the system will require their input. In most cases finances are required and it always requires their approval as they are the ones that give out money to be used in logistics during a discussion of the data tools involved. This means that emerging health needs that may be necessary to add in the system or another system component that are considered redundant cannot be changed based on the donor priorities or availability of funds.

'In case changes are supposed to be done on the data collection tool, we have to do it through our donors. They are the ones who will give us per-diems and organize our transport to the venue where these changes are to be done. This is because changes cannot be made in one facility only. It must be done nationally Said one of the MCH in charge'

Acceptability of a system can be evident on how the stakeholders of the system embrace the system in its usage and compliance with its requirements. Lack of sensitization and enforcement emerged as clear voices as the main impediment to the acceptability of the system. These voices are evident in the quote below where the respondents operating at district level stated that compliance by health facilities in both public and private sectors was good but that private general practices were particularly problematic in under-reporting of MTCT cases and lack of participation in PMTCT meetings as indicated by the verbatim report below. This indicates that a section of the stakeholders of the system in the private sector has not totally embraced the system probably due to lack of sensitization or lack of enforcement.

'Most private facility does not send for representative even if they are called for PMTCT meetings where they also report on MTCT cases. The district PMTCT coordinator meet with public sector health facilities twice a year and make use of the opportunity to give feedback to the reporting units on surveillance data said one of the PMTCT coordinators'

The capability of a data management or a surveillance system to achieve its intended objectives determines the *positive predictive value* of the system. The intention of the EID data management system is to ensure that all MTCT cases are communicated to both the patient and NASCOP in a timely manner. The key emerging voice as to the main impediment to the positive predictive value is the availability of funds to employ defaulter tracing. These defaulter tracers are the ones responsible for tracing all the MTCT cases to ensure that the results are communicated in a timely manner. This is evident in the verbatim below where the respondents indicated that variable amounts of time are expended on following up on reported cases of MTCT and is largely dependent on human resource, and financial resource availability to assist with these investigations.

'It depends on whether a defaulter tracer is there or not. Most of the time those defaulter tracers are employed by the partners said one of the nurses in the facility'.

The composition of various stakeholders meeting and the degree to which the data management system reflects the actual need of what is supposed to be captured reflects the *representativeness* of that system. During analysis misrepresentation and lack of sensitization on the system usage emerged as clear voices impeding the representativeness of the system. In the verbatim below it is important to note that while there is a national consensus on the need to capture all the data elements indicated by the system developer, the importance is yet to be communicated to the staff at the facility as indicated by their question as to why some of the data elements are being captured.

'most of the time we are being told that these data is being used to change the policy, but us we have never seen the policy change or how they can be used to change policy. We sincerely do not know why some of the data elements are being captured. Said one of the PMTCT coordinators'.

Stability of a data management system is the ability of the system to operate flawlessly with minimal interruption that cannot interfere with it in fulfilling its objectives in a timely manner. During analysis, erratic kits supply, shortage of staffing, and equipment breakdown were identified as among the prominent voices that impede the stability of the EID data management system. This is indicated in the verbatim below where respondents indicated that there had been a number of times there is system downtime at the central testing laboratory thus impeding real-time data entry thus weakening the system.

'Sometimes the system goes down at the central testing lab due to kit shortage or equipment downtime or when one of our staff goes on leave since we are few causing a backlog of data entry. This means a lot of time has to be taken when clearing the backlog as reported by one of the laboratory staff'.

4. Discussions

4.1. Accuracy and Completeness

Since the inception of EID testing in a government facility in 2007, the PMTCT health care system in Kenya has been focused on achieving comprehensive and equitable service delivery to the entire HIV Positive pregnant mother population of the country [16]. Similarly the national EID data management system is rooted in a disjointed health service where health information systems have functioned sub-optimally and in some places not at all [17]. The Department of Health of Kisumu East and West district faces the same challenges and facets of its operations are still in a transitional phase. Despite these acknowledged constraints the performance of the department's programmes and routine data management systems needs to be monitored and evaluated [18]. The enthusiasm with which stakeholders participated in and endorsed this evaluation is an indication of willingness to assess and improve EID data management system.

Probably what went wrong is the inception of the EID data management system that makes the data be inaccurate and incomplete is the transition from the fragmented system to the integrated system when all the PMTCT elements were supposed to be grouped together [19]. Davis, Rob in the bulletin "What Is the Difference between Data Validity and Data Integrity?" explains that failure to adequately evaluate the clinical workflows and information need to be associated with providing care and lack of planning during and after transitioning to a new system will result in a fall back to paper, thereby jeopardizing the success of the new data management system [20].

Davis goes and explains that practices implementing new data management system should create a program to promote and monitor data quality after go live [20]. Developing a data quality program requires ensuring data management practice processes are standardized and procedures are uniform. Clear and concise documentation guidelines and training on responsibilities and expectations are imperative and must take into account regulatory, governmental, and accreditation standards (where applicable) [1]. Practices must monitor and audit data for compliance with their data quality programs on both the back end and the front end to ensure accuracy and overall compliance throughout the conversion [11]. They will need to decide what data to clean up, how far back to go, and how long the transition will take. This is what was probably not done considering that even in the same facility set up; the HIV exposed infant register at the MCH while the data details are mostly captured while at the Laboratory.

The public health importance of data elements included in the list of MTCT cases should be reviewed periodically [10]. The current list of data elements in Kenya has remained relatively constant since the regulations for HIV care were submitted by NASCOP to the ministry of health as a policy paper in 2008 [3]. The Ministry of Health has documented the objectives and purposes of the data management system but the absence of documents to that effect at provincial and district levels indicates that that information flow to other levels of the health care system is lacking [21]. Similarly there appears to be some confusion regarding the usage of EID data and MTCT reporting as evidenced in our interviews. The data elements collected on PMTCT cases in Kisumu East and West could be used for many purposes but at present the resulting information is used predominantly for procurement of reagents and consumables and resource allocation for PMTCT activities within the district and to fulfil mandatory requirements. This lack of feedback on the usefulness of complete data filling is what probably is causing these facilities not to complete the data form as they are not seeing the importance of all the information.

The rigidity and "discipline" imposed by the managers for target achievement may be understood as an indication of what Streefland calls, metaphorically, "a military organization model" which so often emerges in PMTCT and EID programs [22]. Further errors in the facility reporting

might be added due to lack of motivation of the health personnel, lack of feedback, no concern for quality information and no cross-checking mechanism. The system in general, as it is designed, invites "data cooking" as well as the lack of interest in supporting practices such as record keeping and data use [22]. For managers the major concern is to achieve the target; therefore, the information system is seen as an "upward system" and not as a system that may support their own work. A common perception is that to improve accuracy and timeliness of data, redesigning the forms and data collection procedures constitutes the main solution [23]. Using this approach, implementing a register book at the facility level to ensure record keeping, could be a suggestion. However, we believe that the most important aspect is to relate information needs to interventions with a focus on how information generated could be used and influence local decisions. Some experiences, for example in Kyrgyzstan and in South Africa, showed improved data quality by giving health workers the basic skills to monitor their own work, leading to a sense of ownership of the generated information [8].

4.2. Best Practices

Different approaches can be used to improve the support mechanisms, for example, increasing the quality of the supervision visits regarding the quality of data from the tally sheet, as well as providing an adequate feedback mechanism to the producers of data at the remote sites. On the other hand, supervision visits could include a more comprehensive data analysis on EID data elements that are required to be filled. It could be used as a way to do on-job training on basic concepts and monitoring indicators, a strategy used in some countries producing satisfactory results. The "eyeballing" approach (a quick look at the forms), the 3C's approach (completeness, correctness and consistency) could also be promoted as the first step towards data quality improvement, and could be an essential part of health workers at the remote sites [23].

All of the attributes of surveillance systems assessed in this study vary in their relative importance depending on the objectives of the system. The attributes may even detract from one another, for example improved timeliness may compromise completeness if data are cumbersome to obtain. This research revealed contradictory views about the usefulness of the EID data management system. The PMTCT coordinators cited that the system was useful in stimulating changes in public health policy in response to surveillance data on MTCT outbreaks. Some of the private general practitioners surveyed, however indicated that they felt there was no purpose in reporting cases as they perceived that there was little public health action arising from the notifications. Feedback sessions are held twice a year within the public sector in Kisumu East and West district between the provincial and local health departments and public hospitals and primary health care facilities. Such interaction is lacking between the public and private sectors and this may result in a lack of communication.

The organisational structure of the EID data management system and the procedures for data management do not conform to ideal routine health information system design. There is potential for duplication of data management because of variations in reporting from the facility, through the district, to provincial and eventually to NASCOP. Reporting procedures vary within facilities. In some public hospitals, laboratories report EID data elements directly to DASCO who pursue the data elements further whereas other facilities rely on med superintendent alerting the DASCO or other designated parties responsible for EID and PMTCT. In the private sector the onus is on the practitioner alone to report the data elements. The human resource structure of the system incorporating provincial and local authority personnel is in a phase of organizational change to form a unified public service. On an interview with key stakeholders in the system it appears that cooperation between provincial and local authority personnel is very good. Cooperative PMTCT and EID teams have been formed in each district and the workload is divided amongst the participants to ensure the smooth success of the program.

There is already a reasonable degree of integration of a number of vertical components at provincial and district EID and PMTCT levels. The expanded program on immunizations and the EID data management system operate through the same departments – this enables the local use of information to monitor the impact of immunizations and PMTCT interventions. The EID data management system in Kisumu East and West district demonstrates reasonable flexibility both in terms of data elements captured and the data submission process. Data element may be added to the reporting list on a local level through communication with reporting units although nationally data element change require a larger PMTCT and EID stakeholder working group. The degree of flexibility of the data submission process will be tested through the recent introduction of electronic data submission from the district level.

The quality of EID data management in Kisumu East and West district is a particular challenge. Interviews with key stakeholders in the system suggest that there is skepticism as to the validity of the data. The example of the incompleteness of reporting of mothers PMTCT intervention that showed a lack of information on this particular data element of up to 42% while the infant breastfeeding status of the infant and infant prophylaxis remained at 77% and 38% respectively. The districts should be monitoring compliance with data management on a checklist of reporting units but due to the fact that there has been no dedicated data evaluation officer at the unit since the inception of the program in 2007, this kind of detailed monitoring is low on the priority list after the more pressing responsibilities of the limited staff complement. Only a few archived data on EID data elements were available at the provincial office but from 2007 compared to the data submitted to NASCOP, and the quality of the data is uncertain. It is not possible retrospectively to determine delays in reporting, the number of repeat testing, or factors leading to seroconversion of infants on PMTCT

without this critical information but the incorporation of recording of these dates, similar patient identifier during repeat testing in the EID data management system would be valuable for prospective analysis.

The positive predictive value of EID data management system has a direct bearing on the resources expended on following up reported cases of MTCT [21]. In a system based on clinical suspicion rather than laboratory results there is the potential for high numbers of false positive reports unless the occurrence of the condition is high and diagnostic features are specific. The EID data management survey respondents all indicated a high workload with a limited staff capacity further emphasizing the importance of reliable notifications.

Subsequent to the completion of this study the data management system process was changed to allow for electronic submission of data from the district level through to provincial and national health departments. This was done through the incorporation of the reporting of data elements into the existing electronic routine health information system. Electronic data entry should ease legibility but basic computer skills will be required. However, this method of reporting does not guarantee that accurate data will be entered. This research provides a baseline against which the performance of the new process can be measured. For areas that do not have access to computers, the data will be entered at the central Labs for submission to the testing lab then eventually to NASCOP.

Provincial and District PMTCT and EID coordinators have repeatedly cited compliance in data elements reporting amongst private practitioners as a major challenge to effective surveillance. This situation is not unique to Kisumu East and West district or even to South Africa. The primary health care provider survey conducted between September and the first week of December 2011 yielded an 18.4% response rate. There may have been non-response bias where practitioners not participating in interviews were systematically different from respondents in that their practices were busier. Given the median number of patients seen by respondents as 20 per day this may confirm that the practices are less busy than average although there is little published on relevant averages within Kenya

In conclusion, this evaluation has revealed strengths and weaknesses in the EID data management system in Kisumu East and West districts in Kisumu County. The personnel working in the data management system appear to be dedicated, informed and conscientious. However, this research suggests that there is a suboptimal use of the information for local action in certain areas. The completeness and accuracy of notification data, as demonstrated in data elements capture, is insufficient to gauge a true picture of the burden of PMTCT in the two districts. Identification of new and emerging health threats will likely be through routes other than the EID data management system. This assessment serves to enlighten policy-makers on the current state of the EID data management system in Kisumu East and West districts. It highlights areas in which improvements can be made as well as reinforcing successful practices. The EID data management system forms an integral part of the health system and public health planning and implementation.

References

[1] Kohn, L. T., Corrigan, J. M. & Molla, S. (2000). To the Err is Human: Building a Safer Health System. National Academy Press. Wshington DC.

[2] Shaller, D. & Consulting, S. (2007). Patient-Centered Care: What Does It Take? *Commonwealthfund* Pub. no. 1067. http://www.commonwealthfund.org/usr_doc/Shaller_patient-centeredcarewhatdoesittake_1067.pdf?section=4039 (Accessed on: 22nd December 2015).

[3] Luoma, M. Doherty, J. Muchiri, S., Barasa, T., Hofler, K., Maniscalco, L., Kirika, R., Ouma, C., & Maundu, J. (2010). *Kenya Health System Assessment 2010*. Bethesda, MD: Health Systems 20/20 project, Abt Associates Inc.

[4] Purkis B, Practice MDM., Morris, G. Associate, S., Afzal, S. (2012). Master Data Management within HIE Infrastructures: A focus on master patient indexing approaches. Audacious Inquiry. 2012.

[5] Odeny, T. A., Penner, J., Lewis-Kulzer, J, Leslie H. H, Shade, S. B., Adero, W., Kioko, J., Cohen, C. R., & Bukusi, E. A. (2013). Integration of HIV Care with Primary Health Care Services: Effect on Patient Satisfaction and Stigma in Rural Kenya. *AIDS Res Treat*; 2013: 485715. PMID: 23738055; PMCID: PMC3664481.

[6] Secretary's Advisory Committee (2008). *National Health Promotion and Disease Prevention Objectives for 2020: Recommendations for the Framework and Format of Healthy People 2020*. Office of Disease Prevention & Health Promotion, U. S. Department of Health. http://www.healthypeople.gov/2010/hp2020/advisory/default.asp (Accessed on 20.03.2016)

[7] Mbor-Ngacha, D and Shaffer, N (2010). Prevention of Mother-to-Child Transmission of HIV (PMTCT) and Paediatric HIV Care and Treatment. *Joint IATT Technical Review mission Report*. Ministry of Health Kenya.

[8] Richardson, J., Robertson, I. K. & Wildman, J. *A (2001). Critique of the World Health Organisation's Evaluation of Health System Performance*. WHO Working paper 125 8–16.

[9] NASCOP *National Communication strategy*.

[10] NASCOP *National EMTCT communication strategy*. 3–5 (2012).

[11] Breu, R., Sztipanovits, J. & Ammenwerth, E. (2008). Model-based design of trustworthy health information systems. *Methods of information in medicine* 47, 389–91.

[12] PEPFER (2012). Guidance for Orphans and Vulnerable Children Programming. http://www.pepfar.gov/documents/organization/195702.pdf (Accessed on 21.01.2016).

[13] Kasiulevičius, V., Šapoka, V. & Filipavičiūtė, R. Sample size calculation in epidemiological studies. *Sample size calculation i* 7, 225–231 (2006).

[14] Hoppanbrowers, J., Jan van den Heuvel, W., Hoppanbrowers, S., Weigand, H & de Troyer, O. (1996). The Grammalizer: A CASE Tool Based on Textual Analysis. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.5950&rep=rep1&type=pdf (Accessed on 12.12. 2016).

[15] Cole, C; Israel, E; Ladha, S; Lulu, K; Mathew, R; Mwakangalu, D; Nouga, A (2013. *Strengthening Community and Health Systems for Quality PMTCT: Applications in Kenya, Nigeria, South Africa, and Ethiop*ia. Pathfinder International, 9 Galen Street, Watertown, MA, 02472, USA

[16] Metsch, R. L., Feaster, J. D., Gooden, L., Schackman, B. R., Matheson, T., Das, MMD. *et al*. (2013). Effect of Risk-Reduction Counseling With Rapid HIV Testing on Risk of Acquiring Sexually Transmitted Infections. *The Journal of the American Medical Association*, 310 (16): 1701-1710, doi: 10.1001/jama.2013.280034

[17] Azcoaga-Lorenzo, A., Ferreira, C., Alvarez, A. P., Palma, P. Velilla, E & delAm, J. (2011) Effectiveness of a PMTCT programme in rural Western Kenya. *AIDS Care*, 23 (3): 274-280, DOI: 10.1080/09540121.2010.507750.

[18] Dillabaugh, L. L., Kulzer, L. J., Owuor, K., Ndege, V., Oyanga, A., Ngugi, E., *et al.* (2012). Towards Elimination of Mother-to-Child Transmission of HIV: The Impact of a Rapid Results Initiative in Nyanza Province, Kenya. *AIDS Research and Treatment*, Volume 2012 (2012), Article ID 602120, 6 pages, http://dx.doi.org/10.1155/2012/602120

[19] Tshibumbu, D. Desire (2006). *Factors influencing men's involvement in prevention of mother to child transmissi on (PMTCT) of hiv programmes in the mambwe district, Zambia*. A dissertsation Submitted in Partial Fulfilment of The Requirements for The Degree of Master of Public Health at the University of South Africa. (Unpublished)

[20] Greyling, C., Kean, S., Perry, F., & Sengstmann, S. (2012). Community Prevention of Mother-to-Child Transmission of HIV (c-PMTCT). Global Health and WASH. World Vision International, the Square, Stockley Park, Uxbridge, Middlesex, UB11, 1FW, United, Kingdom

[21] Ferguson, L., Grant, A. D. Ong'ech, O. D., Vusha, S., Deborah Watson-Jones, D. & Ross, A. D. (2012). Prevention of mother-to-child transmission of HIV: assessing the accuracy of routinely collected data on maternal antiretroviral prophylaxis coverage in Kenya. *Sexually transmitted infections*; 88 (2): 120-4.

[22] Fleischman, J. (2011). *Lessons from Kenya for the Global Health Initiative Lessons from Kenya for the Global Health Initiative*. Center for Strategic and International Studies, 1800 K Street, NW, Washington, DC, 20006.