



FIELD TESTING OF A SIMPLE ODK APPLICATION FOR DOUBLE DATA ENTRY FOR COMMUNITY-BASED HEALTHCARE RESEARCH AND DISEASE SURVEILLANCE

Iwara I. Arikpo

Department of Computer Science, and
HDSS UNICAL, Directorate of Research
University of Calabar
Calabar, Nigeria

Anthony T. Okoro

Department of Computer Science, and
HDSS UNICAL, Directorate of Research
University of Calabar
Calabar, Nigeria

Ememobong Aquaisua

HDSS UNICAL, Directorate of Research
University of Calabar
Calabar, Nigeria

Martin M. Meremikwu

Department of Paediatrics, College of Medicine, and
HDSS UNICAL, Directorate of Research
University of Calabar
Calabar, Nigeria

Abstract: Data quality plays a vital role in the reliability of data for planning and decision making. The methods used for data collection and entry further heightens the concern for data quality. This paper addresses the techniques of double data entry as an efficient and simplified approach designed to improve the quality of paper-based records.

Data from implementation of the operations research component of a larger health intervention project in Abia State, South-east, Nigeria, was used for the implementation. Paper-based data were entered independently by two data entry clerks with unique identifiers (IDs) using ODK application. The data was then exported in .csv format into Microsoft Excel application and compared for discordant entries.

The algorithm auto-compared all records by the data entry clerks and returned zero and non-zero values for all concordant and discordant entries respectively. This allowed for easy spot checks on the questionnaires and subsequent correction of the erroneous entries.

Double data entry is efficient, cost effective and robust in achieving high data quality with paper-based records.

Keywords: data quality; double data entry; data cleaning; paper-based records; data entry errors; paper forms

I. INTRODUCTION

Generally, researches are carried out to discover truths, validate hypothesis, increase knowledge, and foster better and evidence-based solutions to established problems of interest. The appropriateness research design and quality of data collection process play vital role in the outcome, credibility and acceptance of research results. As opined by Krishnankutty et al. (2012), high-quality data are expected to be “absolutely accurate and suitable for statistical analysis; meeting the protocol-specified parameters and comply with the protocol requirements.” This highlights the significance of good data in every scientific research as outputs from the analysis of the data are used by scientists, governments, organizations, individuals, etc. for making projections, planning and other decision-making activities.

To minimize the introduction of errors and ensure that quality of research data is guaranteed, different approaches are often adopted by researchers at various stages of the study. Integrity measures such as controls, built-in logics and validation checks are frequently applied in data collection and entry processes to minimize the occurrence of errors. These data assurance processes are nowadays more frequently applied used in electronic data collection involving systems like personal digital assistants (PDAs),

tablets or smartphones. Notwithstanding the integrity measures, errors and omissions are still common with data entry activities, certainly due to known inevitable factors such as speed of entry, fatigue, age and level of experience of the data clerks (Scott, Thompson, Wright-Thomas, Xu, & Barchard, 2008). These data entry errors can sometimes be so severe as to invalidate inferences and conclusion from results of the study. Double data entry is one step that has been shown to reduce data entry errors and improve the quality of research outputs (Coleman Data Solutions, 2014).

Double data entry (also known as two-pass verification) is a traditional data quality control measure that requires same set of data to be entered more than once in order to make comparison (by computer applications), reconcile and correct dissimilarities arising from key-punch errors and missed values in the entry (Gregg, 2008). Different forms of double data entry exist. One is applied when a data entry clerk enters a set of data into the electronic system (i.e. computers, mobile devices, PDAs, etc.) and repeats the entry for the same set of data. The other, as used in this research, requires the same set of data to be entered by two different data entry clerks on the assumption that there is limited possibility of the two clerks making the same data entry mistake on a dataset. The data clerks are usually assigned systems and unique codes for easy identification and cross-matching of entries.

The central idea for the rigorous processes of double data entry is to have datasets that fulfill the characteristics of data quality requirements such as correctness, consistency, completeness and currency of the data. These data quality control and assurance characteristics are easily achieved if a good standard of verification and validation is used during the data entry. Double data entry in combination with built-in validation and logic constraints at the point-of-entry promotes the attainment of these standards, since errors arising mostly from discordant entries and missing data are detected and corrected.

Double data entry technique is a common principle and highly used approach in the conduct of many surveys to verify the correctness of survey data. For instance, surveys such as Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) are conducted with the use of double data entry technique as part of best practice approach in processing the survey data (IHSN, 2017). The verification process often adopted in these surveys entails independent entry of research data into separate files by data entry clerks and then comparison of the files to identify and reconcile inconsistencies in the entry, either by using a standalone application or application with built-in features for double entry verification.

Over the years, different applications have been developed to handle double data entry operations. In 1992, a double entry verification routine was developed by McDonald to compare the uniqueness of data sets at the end of entry of the observations. Though the program had such limitations as its inability to identify duplicates and/or missing entries in the data sets, it was effective at the time. A similar application that supports verification of entered data was created by Boyle and Brinsfield using SAS/FSP™ software. The system required the entry to be done in sequence. At the completion of entry, a Screen Control Language (SCL) was used to compare the entries, report and correct discrepancies (Shaffer & Groninger, 1995). The need to enforce data quality through double entry has continued to evolve. Shaffer and Groninger (1995) developed an application for validating data entry using double-entry techniques.

Double data entry facilities have also been automated into some data management applications to promote data integrity. An example is OpenClinica (Open-source software for Clinical research) - one of the most widely used web-based healthcare management software with built-in facilities to support double data entry. Though an optional feature in the OpenClinica system, double data entry is done by keying completed paper-based Case Report Forms (CRFs) into the system and then re-entering of the same form by another user or the first user, provided the latter is done after a minimum 12 hours (OpenClinica, 2017). The process enables the application to compare and flag differences in the entries.

CSPRO (Census and Survey Processing System) is another public domain software application used by many organizations and individuals for data entry, management and analysis of census and survey data (Census.gov, 2017). A key feature of CSPRO is its support for double data entry. The 'compare data' tool of the software allows for the comparison of contents of files which were independently

entered into the system. Once an entry is made, the compare data tool verifies the entry with the earlier data in the data file and activates prompt for next command if the entries are concordant. In cases of mismatches, a message is flagged indicating the dissimilarity. The data are then re-entered. EpiData (EpiData, 2019) and RedCap (REDCap, 2019) are among many other software tools that have features for double data entry.

Double data entry technique is usually a key requirement used by many organizations and governments for the assurance of data quality and enhance the credibility of disease surveillance records and research output. In 2006 for instance, the Rwandan National Institute of Statistics included double data entry as one of the key data processing specifications for the Integrated Household Living Conditions survey 2005-2006 to perform independent verification of entered data (NISR, 2016). In 2014, Ndume, Nkansah-Gyekye and Ko developed an e-health software system that can be used to evaluate the quality of paper-based data. The system's algorithm was to identify and remove duplicates and spelling mistakes. Though they recorded an improved data quality through this algorithm, the issue of correcting the erroneous data without re-entering the data from scratch was not addressed.

Double data entry is indeed an effective means of data verification and error reduction during entry, the need for additional resources such as personnel, time and finance are common challenges (Gisela, Birgit, Gabriele, & Stephan, 2005). These limitations should be taken into consideration during the planning phase of any project requiring data collection and entry. Loss and mix-up of questionnaires by the data clerks have been reported a limitation double data entry especially in settings where document management processes are suboptimal. Some of the known computer-based applications with facility for double data entry require that the first enterer keys in all the data before the second enterer. It is at the point of data entry by the second enterer that the system performs a matching process of the data being keyed in with that already entered.

We the development and testing of a simple ODK application that automates quality checks and controls and capacity for concurrent, independent double data entry. We sought to address such known limitations of double data entry as timeliness and need for additional personnel skills while ensuring that the effectiveness of double data entry regarding data quality is achieved.

II. METHODOLOGY

A. Setting

The research setting under which the proposed simplified approach was tested is an iCCM (integrated Community Case Management) programme, implemented in 15 local government areas (districts) of Abia State, South-east Nigeria. The goal of the project was to reduce child mortality from pneumonia, diarrhoea, malaria and malnutrition among children ages 2-59 months, thereby accelerating achievement of health-related millennium development goals (MDGs) in the State (Region). The authors were involved in the operations research (OR) component of the project. In the iCCM programme, frontline health workers community health workers (CHWs) who administer healthcare to children under 5 years of age,

at the community level, with a target population of 202,198 children.

Under the operations research component, the research sites were grouped into Intervention and Control arms, respectively. A normal iCCM programme was implemented in the latter, while in the Intervention Arm, the iCCM programme was implemented with an additional component, called Peer Group Supervision (PGS). This component introduced a horizontal layer of supervision, where CHWs supervise each other by holding monthly meetings to review cases and patient registers of each other, rendering corrective advice where necessary. These meetings were alternated among the CHWs in each cluster to ensure that each CHW has an opportunity to preside. Various data sets (Table 1) were collected at baseline and endline points of the operations research component, which was entered by double data entry technique. Questionnaires used for the data collection were standardized and approved by the WHO. Data collection was done by trained Nurses/Midwives under the supervision of the Author Team.

Table I. Type of information collected from the operations research component

Questionnaire	Data collected
Direct observation checklist	Comprehensive patient's data for children up to five (5) years.
Re-examination	Data on the assessment of CHW's decisions and classification based on a child's symptoms.
Exit interview	Collects information from a caregiver on his/her perceptions about the services rendered by a CHW.
Equipment, supplies & support checklist	Collects information on viability of equipment and working tools with a CHW including stock-out of drugs and supervisory visits.
CHW Sociodemographic & background information	Collection of information on the sociodemographic and background of CHWs, such as age, education, work experience, etc.
Case scenarios	Collects information on case scenarios on the three childhood diseases demonstrated by CHWs, such as, pneumonia, diarrhoea, and malaria.
CHWs motivation	Collects data on the level of motivation of CHWs on their services to community.

B. Procedure for handling double data entry

Data was collected with paper questionnaires by the Nurse/Midwives. Data entry was done with mobile devices running the Android operating system installed with

ODKCollect application (Open Data Kit, 2017). Each questionnaire was entered by a pair of properly identified DEC's (data entry clerks) and transmitted to a GoogleCloud Server (Google Cloud Platform, 2016) running the ODKAggregate. The data is then downloaded in CSV (comma separated values) format and saved in a standard spreadsheet, like Microsoft Excel. The steps to checking for correctness from the doubly-entered data are laid out in the following pseudocode, while the corresponding flowchart is shown in Fig. 1:

- Sort on QuestionnaireID and DEC_ID. This will create pairs of QuestionnaireID, but single entry for each pair of the former.
- Create a temporary column called Temp_Series and create a numerical series on it (e.g., 1, 2, 3, 4, 5, 6, 7, ..., n).
- Create a temporary column called Series_MOD to compute modulus on the series using the formula: $=MOD(Temp_Series, 2)$ [e.g., $=MOD(E2, 2)$]. From the outcome of the formula, '1' represents the odd record numbers, '0' represents the even record numbers. Each of these will respectively represent DEC1 and DEC2.
- Sort on Series_MOD in descending order (largest to lowest).
- Under the DEC_ID column, fill all cells corresponding to '1' in the Series_MOD column with 'DEC1', and all cells corresponding to '0' with 'DEC2'.
- Delete columns: Temp_Series and Series_MOD.
- Split the worksheet into DEC1 and DEC2 to contain data entered by each DEC respectively.
- Create a blank worksheet (after DEC2) and call it 'Compare' to compare entries from DEC1 with DEC2.
- Copy the column headings from any of DEC1 or DEC2 into the first row of 'Compare'.
- Create the formula: $IF(EXACT('DEC1'!A2,'DEC2'!A2),0,'DEC1'!A2\&"/\&'DEC2'!A2)$ say in A2.
- Copy the formula into all cells. Note that, cells in the 'Compare' worksheet showing '0' imply that, DEC1 entered exactly the same data item with DEC2, otherwise, the cell will show what was entered by each DEC.

III. RESULTS AND DISCUSSION

Fig. 2 depicts the data table as entered by one data clerk. Similar data table applies to the second data clerk with only the DEC_ID code being different. At this stage, the dataset looks complete. Only minimal errors such as missing values and typos can be seen easily.

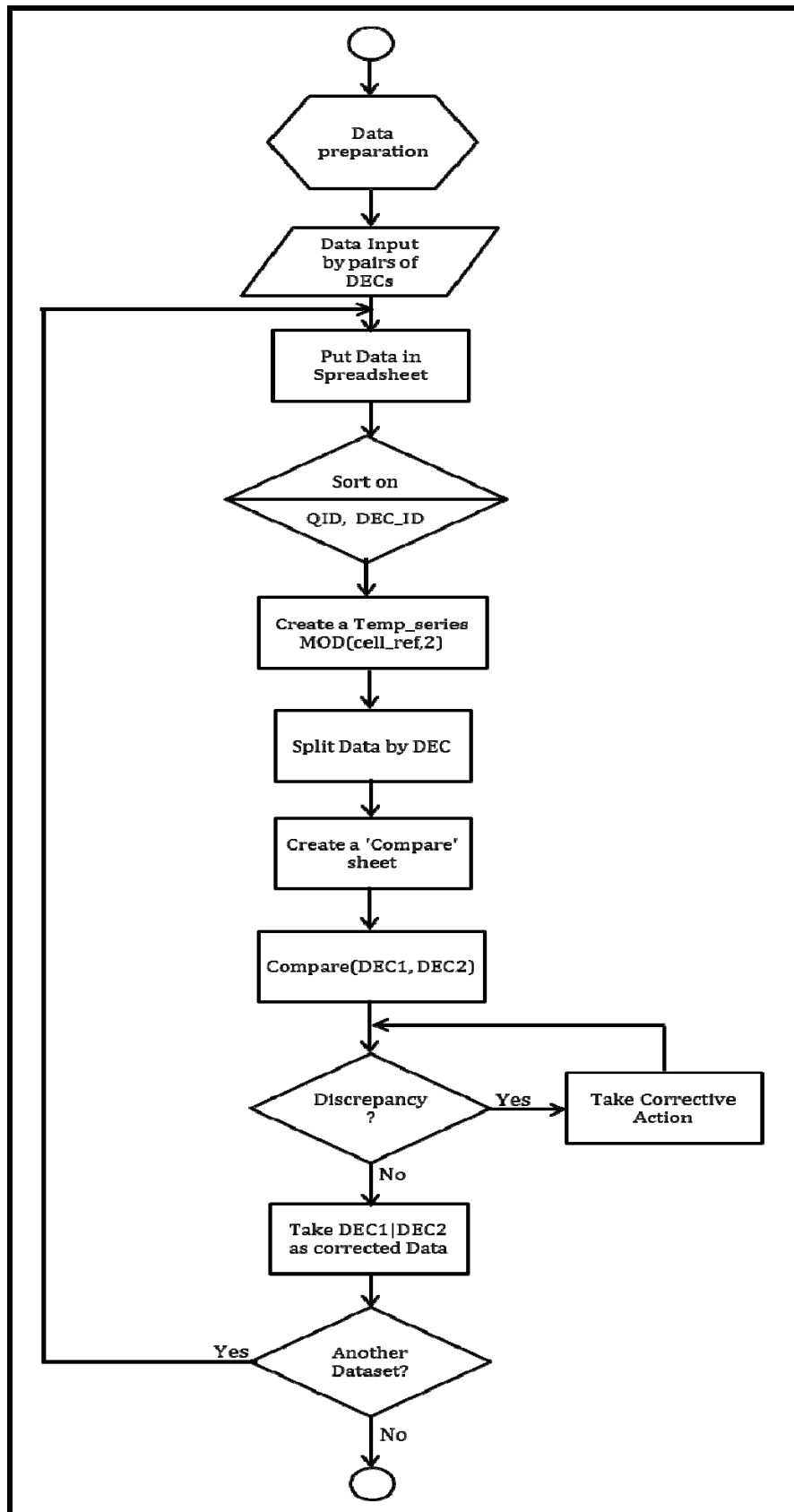


Figure 1. System flowchart for the double data entry process

	B	C	D	E	F	G	H	I	J	K	
1	DEC_ID	Date_of_interview	LGA_code	StartTime	EndTime	Evaluator_Code	CORP_Code	SD1_CORP_age	SD2_CORP_gender	SD3_Highest_level_of_educatn	SD3_specif
2	DEC1	Thu Feb 23 00:00:00 UTC 2017	11	9:36:00	12:00:00	CE1104	CE1135		56 Female	Secondary_School	
3	DEC1	Wed Feb 22 00:00:00 UTC 2017	11	10:37:00	9:40:00	CE1103	CE1116		42 Female	Tertiary	
4	DEC1	Thu Feb 23 00:00:00 UTC 2017	11	10:05:00	10:08:00	CE1103	CE1134		31 Female	Secondary_School	
5	DEC1	Wed Feb 22 00:00:00 UTC 2017	11	2:33:00	2:35:00	CE1103	CE1120		32 Female	Secondary_School	
6	DEC1	Wed Feb 22 00:00:00 UTC 2017	11	1:31:00	1:34:00	CE1103	CE1115		22 Female	Secondary_School	
7	DEC1	Wed Feb 22 00:00:00 UTC 2017	11	0:48:00	23:00:00	CE1104	CE1129		45 Female	Secondary_School	
8	DEC1	Wed Feb 22 00:00:00 UTC 2017	11	10:14:00	10:16:00	CE1104	CE1128		47 Female	Secondary_School	
9	DEC1	Tue Feb 21 00:00:00 UTC 2017	11	11:02:00	11:03:00	CE1104	CE1114		45 Female	Secondary_School	
10	DEC1	Tue Feb 21 00:00:00 UTC 2017	11	2:46:00	2:49:00	CE1104	CE1133		35 Female	Secondary_School	
11	DEC1	Mon Feb 20 00:00:00 UTC 2017	11	12:09:00	12:17:00	CE1104	CE1117		50 Female	Secondary_School	
12	DEC1	Tue Feb 21 00:00:00 UTC 2017	16	13:30:00	14:00:00	CE1605	CE1612		43 Female	Secondary_School	
13	DEC1	Tue Feb 21 00:00:00 UTC 2017	16	7:20:00	8:10:00	CE1605	CE1614		45 Female	Secondary_School	
14	DEC1	Tue Feb 21 00:00:00 UTC 2017	16	14:00:00	15:20:00	CE1605	CE1613		42 Female	Secondary_School	
15	DEC1	Wed Mar 01 00:00:00 UTC 2017	16	9:30:00	12:40:00	CE1605	CE1603			Other_education	OND
16	DEC1	Wed Mar 01 00:00:00 UTC 2017	16	12:58:00	13:00:00	CE1605	CE1609		42 Female	Other_education	TCTWO
17	DEC1	Mon Feb 27 00:00:00 UTC 2017	16	9:40:00	14:10:00	CE1605	CE1604		29 Female	Other_education	NCE
18	DEC1	Sun Feb 26 00:00:00 UTC 2017	16	10:30:00	12:05:00	CE1605	CE1610		52 Female	Other_education	OND
19	DEC1	Fri Feb 24 00:00:00 UTC 2017	16	13:50:00	14:00:00	CE1605	CE1608		41 Female	Secondary_School	
20	DEC1	Sun Feb 26 00:00:00 UTC 2017	16	14:30:00	15:20:00	CE1605	CE1616		42 Female	Secondary_School	
21	DEC1	Sat Feb 25 00:00:00 UTC 2017	16	13:50:00	14:45:00	CE1605	CE1602		39 Female	Other_education	NCE
22	DEC1	Tue Feb 28 00:00:00 UTC 2017	16	8:30:00	11:10:00	CE1605	CE1601		52 Female	Tertiary	
23	DEC1	Mon Feb 27 00:00:00 UTC 2017	16	14:23:00	16:30:00	CE1605	CE1605		43 Female	Secondary_School	

Figure 2. Data before applying the formula

Applying the algorithm as shown in Fig. 1 will compare both data tables; reporting discordant entries as seen in Fig. 3. This result identifies all data points where both data clerks keyed in the exact values by returning zero in the Compare table. Similarly, the non-zero values are returned for data points where both clerks keyed in different values. The forward slash '/' symbol separates the values entered by the first and second data clerks, respectively. In the result shown in Fig. 3, cell J45 shows that data clerk 1 keyed in 'Female' for SD2_CHW_gender while the variable was left blank by data clerk two for the same record. Similar cases of mismatching entries are shown in cells H52, H54, O64, P64, among others. Regardless of the dataset size and number of variables in the dataset, the algorithm compares the records automatically and generates the reports as seen in Fig. 3.

It would be wrong at this stage to assume and attribute the error in the entry to a data entry clerk without retrieving the paper questionnaire to check for the exact value written. Once the questionnaire is retrieved and the correct value found for the mismatched variable, the worksheet containing the wrongly inputted value is loaded and the correct value entered. The process maybe tiring for inexperienced data clerk who would retrieve questionnaires, correct detected error for a variable, return the questionnaire back to archive and then retrieve again at a later time for another variable with mismatching entries. The easy and fast rule which has proven to be highly efficient and time-saving is to retrieve a questionnaire of interest, correct error in hand and then walk through the questionnaire to the end of the record in the Compare sheet using the record ID before re-archiving. At any time, an error is found, the worksheet with erroneous value is loaded and the correct value entered.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Questionr	DEC_ID	Date_of_interview	LGA_code	StartTime	EndTime	Evaluator	CORP_Coc	SD1_CORF	SD2_CORF	SD3_Highest_level_o	SD3_speci	SD4_CORF	SD4_othe	SD5I_Mof	SD5II_YEA
44	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	0	DEC1/DEC	0	0	0	0	0	0	0	Female/	0	0	0	0	0	0
46	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	0	DEC1/DEC	0	0	0	0	0	CE0428/CE	0	0	0	0	0	0	0	0
53	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54	0	DEC1/DEC	0	0	0	0	0	CE0409/CE	0	0	0	0	0	0	0	0
55	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57	0	DEC1/DEC	Thu Feb 02 00:00:00	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	DEC1/DEC	Thu Mar 02 00:00:00	0	0.949305555	0.951388888888	0	0	0	0	0	0	0	0	0	0
59	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	DEC1/DEC	Thu Mar 02 00:00:00	0	0	0	0	0	0	0	0	0	0	0	0	0
61	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	0	DEC1/DEC	0	0	0	0.81597222222	0	0	0	0	0	0	0	0	0	0
63	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
64	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	12/	2014/12
65	0	DEC1/DEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3. The 'Compare' sheet showing mismatches after applying the formula

The process is repeated for all cases where mismatching entries are found. Fig. 4 shows the Compare table with zero values for all data points. This implies that all entries by data clerk 1 and data clerk 2 have been resolved; thereby

validating the correctness of the datasets as collected by the interviewers. At this point, any of the data tables can be used for analysis, since both files are now the same.

DEC_ID	Date_of_interview	LGA_code	StartTime	EndTime	Evaluator_Code	CORP_Code	SD1_CORP_age	SD2_CORP_gender	SD3_Highest_level_of_educatn	SD3_specify_other_education
2	DEC1/DEC2	0	0	0	0	0	0	0	0	0
3	DEC1/DEC2	0	0	0	0	0	0	0	0	0
4	DEC1/DEC2	0	0	0	0	0	0	0	0	0
5	DEC1/DEC2	0	0	0	0	0	0	0	0	0
6	DEC1/DEC2	0	0	0	0	0	0	0	0	0
7	DEC1/DEC2	0	0	0	0	0	0	0	0	0
8	DEC1/DEC2	0	0	0	0	0	0	0	0	0
9	DEC1/DEC2	0	0	0	0	0	0	0	0	0
10	DEC1/DEC2	0	0	0	0	0	0	0	0	0
11	DEC1/DEC2	0	0	0	0	0	0	0	0	0
12	DEC1/DEC2	0	0	0	0	0	0	0	0	0
13	DEC1/DEC2	0	0	0	0	0	0	0	0	0
14	DEC1/DEC2	0	0	0	0	0	0	0	0	0
15	DEC1/DEC2	0	0	0	0	0	0	0	0	0
16	DEC1/DEC2	0	0	0	0	0	0	0	0	0
17	DEC1/DEC2	0	0	0	0	0	0	0	0	0
18	DEC1/DEC2	0	0	0	0	0	0	0	0	0
19	DEC1/DEC2	0	0	0	0	0	0	0	0	0
20	DEC1/DEC2	0	0	0	0	0	0	0	0	0
21	DEC1/DEC2	0	0	0	0	0	0	0	0	0
22	DEC1/DEC2	0	0	0	0	0	0	0	0	0
23	DEC1/DEC2	0	0	0	0	0	0	0	0	0

Figure 4. The 'Compare' sheet after mismatches have been resolved

Current best practice guidelines on data quality assurance discourage the practice of single data entry for data that were collected on paper. Double data entry is the goal standard and minimizes error. Surprisingly, researchers who still adopt single data entry of paper records usually claim to double-check a certain percentage of the entries for errors. Experience has shown that the re-checking of the entries is often not done well. More common practice is the application of some set of quick tests on the variables; to see if the variables contain required values. The algorithm requires the system to flag implausible values. Though this approach might help in correcting some errors, it will not detect data entry errors that fall within the range of plausible data values. The use of double data entry remains the goal standard for dealing with entry of paper-based records. As shown in Fig. 2, errors arising from missing values can be detected by mere scanning of the dataset, even though this becomes impracticable and time consuming for huge datasets with many variable lists. Again, a wrongly inputted value, for instance, the entry of 45 rather than 54 in the SD1_CHW_age variable would be absolutely difficult to detect and correct. Many of such errors are common and bound to occur with single source of data entry.

The Compare tool employed in this paper has demonstrated a free and simplified approach for achieving high quality data through double entry. The algorithm is easy to adopt and implement for all datasets of any kind. All that is required is ability to load the data into Microsoft spreadsheet application, then apply the algorithm as described in the foregoing discussion.

The implementation of this simplified approach is very cost efficient and requires less experience in Information Technology (IT) or software programming skills to implement. There would be lesser need for budgeting huge resources into the use of customized software products for double data entry and cleaning.

IV. CONCLUSION

Application of double data entry technique in data processing is often a contentious issue. While some researchers tend to only promote the enforcement of stringent integrity checks into survey tools, others view double data entry as an added advantage that prevents the

circumvention of data during entry stage. This study demonstrated the effectiveness of applying the latter as an added data quality measure in any research design to strengthen the quality of research data. Regardless of its obvious inability to track errors on the paper questionnaires, double data entry is efficient in identifying discordant entries.

V. ACKNOWLEDGEMENT

This paper is a data science part of the operations research component of a WHO-funded project in Abia State, South-east, Nigeria (Nov 2015 – Jan 2017).

VI. REFERENCES

- [1] Census.gov. (2017). International Programs - Census and Survey Processing System (CSPPro) Documentation. Retrieved from United States Census Bureau: <https://www.census.gov/population/international/software/csp/ro/csprodocs.html>
- [2] Coleman Data Solutions. (2014). Advantages of Double Key Data Entry and Verification. Retrieved from Coleman Data Solutions: <http://www.blog.coleman-data.com/advantages-double-key-data-entry-verification/>
- [3] EPiData. (2019). EpiData Software. Retrieved from <https://www.epidata.dk/index.htm>
- [4] Google Cloud Platform. (2016). Google Cloud Computing. Retrieved January 2015, from Google Cloud Platform: <https://cloud.google.com/>
- [5] Gregg, M. (2008). Field Epidemiology (3rd ed.). New York: Oxford University Press, Inc.
- [6] IHSN. (2017). Finalize collection (data entry). Retrieved from International Household Survey Network: <http://www.ihsn.org/node/111>
- [7] Krishnankutty, B., Bellary, S., Kumar, N., & Moodahadu, L. (2012). Data management in clinical research: An overview. Indian Journal of Pharmacology, 44(2), 168-172. doi:10.4103/0253-7613.93842
- [8] Ndume, V., Nkansah-Gyekye, Y., & Ko, J. (2014). Double Entry Data Capture as an Alternative Solution for Quality of E-Health Records. International Journal of Computer Applications, 93(4), 29-35. doi:10.5120/16204-5490
- [9] NISR. (2016). Data Processing. Retrieved from Rwanda - Integrated Household Living Conditions Survey 2005-2006: <http://microdata.statistics.gov.rw/index.php/catalog/39/dataprrocessing>
- [10] Open Data Kit. (2017). Open Data Kit: Collect. Retrieved 2017, from OpenDataKit: <https://opendatakit.org/use/collect/>

- [11] OpenClinica. (2017). Entering Data for an Event Into CRFs. Retrieved from OpenClinica Reference Guide: <https://docs.openclinica.com/3.1/openclinica-user-guide/submit-data-module-overview/enter-data-event-completing-crfs#content-title-4845>
- [12] REDCap. (2019). REDCap. Retrieved from <https://www.project-redcap.org/>
- [13] Scott, J., Thompson, A., Wright-Thomas, D., Xu, X., & Barchard, K. (2008). Data entry methods: Is double data entry the way to go? Poster presented at the Western Psychological Association Annual Convention. Irvine, CA.
- [14] Shaffer, T. L., & Groninger, N. P. (1995). Double Data Entry for Novice Users. 20th Annual SAS Users Group International Conference (pp. 111-1116). Orlando, FL: SAS Institute, Inc.