

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Steven Russell

Date

Methods in Measuring Surveillance Disease Data Quality, Somalia

By

Steven Russell

MSPH

Emory University

Rollins School of Public Health

Department and Biostatistics

_____ [Chair's Signature]
Paul S. Weiss

_____ [Member's Signature]
George Cotsonis

_____ [Member's Signature]
Curtis Blanton

_____ [Member's Signature]
Susan Cookson

Methods in Measuring Surveillance Disease Data Quality, Somalia

By

Steven Russell

B.S.

James Madison University

2012

Thesis Committee Chair: Paul Weiss, M.S.

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

In partial fulfillment of the requirements for the degree of

Master of Science in Public Health

In Biostatistics

2015

Abstract

Methods in Measuring Surveillance Disease Data Quality, Somalia

By Steven Russell

Introduction:

The ongoing conflict (2011-2015) and famine (2011-2012) in Somalia have presented challenges to collection of surveillance disease data. In order to inform effective intervention decisions in humanitarian emergencies, the data analyst needs a method to assess data quality based on the database alone. The purpose of this study is to develop methods that allow the data analyst to assess the data quality and to develop a standard index of data quality.

Methods:

We scored each facility from 0 to 1 based on 10 individual data quality attributes: proportion of missing weeks, proportion of weeks where counts were all zeros, results of digit preference tests, results of sex ratio tests, proportion of weeks with reporting mistakes, proportion of weeks that did not sum correctly, proportion of duplicate weeks, proportion of weeks with duplicate case counts, and results for 2 different methods of outlier detection. Scores on each attribute were summed and each facility was given an overall score out of 10. A one-way Analysis of Variance (ANOVA) was run to test the differences in data quality scores between the 4 zones of Somalia.

Results:

The overall data quality score for each of our 198 facilities, as well as the facility's score on each quality attribute, were calculated and summarized. Over all facilities, the data quality scores ranged from 0 to 1.418 (mean=.241, median= 0.130, sd =.270), with higher scores indicating more severe data quality issues. On average, facilities in the Southern Zone had the worst data quality scores (mean=.273) while facilities in Somaliland had the best scores (mean=.220). We found no significant differences between the data quality scores in different zones ($F=0.38$, $p=0.77$). Full reports for each facility are provided.

Discussion:

In the future, we hope to further refine the current methods using other surveillance disease datasets. We feel that our data quality index can be an extremely valuable tool in evaluating and improving surveillance systems in humanitarian emergencies. It will allow the data analyst to pick up on unusual patterns that otherwise may have remained undetected and will improve the ability of those on the ground to reduce mortality and morbidity.

Methods in Measuring Surveillance Disease Data Quality, Somalia

By

Steven Russell

B.S.

James Madison University

2012

Thesis Committee Chair: Paul Weiss, M.S.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
In partial fulfillment of the requirements for the degree of
Master of Science in Public Health
In Biostatistics
2015

Acknowledgements

I would like to thank my advisor Paul Weiss for all of the advice and support he has given me on this thesis. I would also like to thank Susan Cookson and Curtis Blanton at the CDC for giving me the incredible opportunity to work on this project and for everything they have done for me over the past year. I also want to give thanks to the people of Somalia and the staff of WHO Somalia for all of their hard work and for permission to use their data. Finally I would like to thank George Cotsonis for taking the time to read over my thesis.

Table of Contents

1. Introduction	1
1.1 Complex Humanitarian Emergencies	1
1.2 Background for Somalia	1
1.3 Surveillance	1
2. Methods	2
2.1 Completeness of the Data	3
2.2 Internal Consistency.....	3
2.3 Duplication.....	4
2.4 Outliers.....	5
2.5 Testing Differences Between Zones	6
3. Results	6
3.1 Southern Zone	6
3.2 Central Zone	8
3.3 Puntland Zone	9
3.4 Somaliland Zone	12
3.5 Overall Comparisons	14
4. Discussion	15
5. Conclusion	16
6. Figures and Tables	17
7. References	41

List of Tables and Figures

Tables

Table 1. Data quality score by facility, Southern Zone, Somalia, 2013	7
Table 2. Data quality score by facility, Central Zone (Facilities 1-30), Somalia, 2013	8
Table 3. Data quality score by facility, Central Zone (Facilities 31-61), Somalia, 2013	9
Table 4. Data quality score by facility, Puntland Zone (Facilities 1-25), Somalia, 2013	10
Table 5. Data quality score by facility, Puntland Zone (Facilities 26-46), Somalia, 2013	11
Table 6. Data quality score by facility, Somaliland Zone (Facilities 1-26), Somalia, 2013	12
Table 7. Data quality score by facility, Somaliland Zone (Facilities 27-54), Somalia, 2013	13
Table 8. Mean, median standard deviation, minimum and maximum scores by zone.....	14
Table 9. Classification of Data Quality Scores.....	14
Table 10. Data quality classification by zone (with percentage of classification per facility).....	15
Table 11. Completeness and Duplication Report, Southern Zone, 2013	20
Table 12. Week Specific Quality Issues, Southern Zone, 2013.....	21
Table 13. Digit Preference Breakdown, Southern Zone, 2013.....	22
Table 14. Completeness and Duplication Report (1), Central Zone, 2013	23
Table 15. Completeness and Duplication Report (2), Central Zone, 2013	24
Table 16. Week Specific Quality Issues (1), Central Zone, 2013.....	25
Table 17. Week Specific Quality Issues (2), Central Zone, 2013.....	26
Table 18. Digit Preference Breakdown (1), Central Zone, 2013.....	27
Table 19. Digit Preference Breakdown (2), Central Zone, 2013.....	28
Table 20. Completeness and Duplication Report (1), Puntland Zone, 2013	29
Table 21. Completeness and Duplication Report (2), Puntland Zone, 2013	30
Table 22. Week Specific Quality Issues (1), Puntland Zone, 2013.....	31
Table 23. Week Specific Quality Issues (2), Puntland Zone, 2013.....	32
Table 24. Digit Preference Breakdown (1), Puntland Zone, 2013.....	33
Table 25. Digit Preference Breakdown (2), Puntland Zone, 2013.....	34
Table 26. Completeness and Duplication Report (1), Somaliland Zone, 2013	35

Table 27. Completeness and Duplication Report (2), Somaliland Zone, 2013	36
Table 28. Week Specific Quality Issues (1), Somaliland Zone, 2013.....	37
Table 29. Week Specific Quality Issues (2), Somaliland Zone, 2013.....	38
Table 30. Digit Preference Breakdown (1), Somaliland Zone, 2013.....	39
Table 31. Digit Preference Breakdown (2), Somaliland Zone, 2013.....	40

Figures

Figure 1. Distribution of scores by zone, Somalia, 2013.....	17
Figure 2. Distribution of scores- Central Zone, Somalia, 2013.....	17
Figure 3. Distribution of scores- Southern Zone, Somalia, 2013.....	18
Figure 4. Distribution of scores- Somaliland Zone, Somalia, 2013.....	18
Figure 5. Distribution of scores- Puntland Zone, Somalia, 2013.....	19

1. Introduction

1.1 Complex Humanitarian Emergencies

The United Nations (UN) has defined a complex humanitarian emergency as “a humanitarian crisis in a country, region or society where there is total or considerable breakdown of authority resulting from internal or external conflict and which requires an international response that goes beyond the mandate or capacity of any single and/or ongoing UN country program.”¹

During humanitarian emergencies, the routine delivery of healthcare is often impeded. Risk factors including mass movement of populations, overcrowded temporary settlements, scarcity of safe water, poor sanitation, and nutritional deficiencies due to food shortages can increase the spread of communicable diseases². Refugees and internally displaced persons are at great risk for morbidity and mortality from these communicable diseases, making effective disease surveillance imperative^{3,4}.

1.2 Background for Somalia

The Somali Civil War is an ongoing conflict that began in 1991 with the fall of former dictator Mohamed Siad Barre⁵. Since that time, Somalia has been without a stable central government and the country has been plagued by lawlessness and warfare. In 2011, the most recent stage of the war began when a joint military operation against Al-Shabaab militants was conducted by the Somali, Kenyan, and Ethiopian militaries^{6,7}.

Beginning in July 2011, severe drought affected the Eastern Horn of Africa, with 2011 being recorded as the driest year since 1995⁸. The UN declared a subsequent famine which lasted until 2012, by which time nearly 260,000 people were estimated to have died either by drought or conflict⁹. The ongoing conflict and the famine presented challenges to collection of disease surveillance and nutrition data.

A number of isolated surveillance systems existed in Southern and Central Zones of Somalia prior to the 2011 drought and famine. These included the Communicable Disease Surveillance and Response system (CSR), the Integrated Disease Surveillance and Response system (IDSR), and disease specific surveillance systems for acute flaccid paralysis, malaria, and measles. The World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) partnered together to assess the CSR and IDSR systems in 2011 and subsequently created one standardized system for all of Somalia¹⁰.

1.3 Surveillance

With the increased risks of morbidity and mortality during humanitarian emergencies, communicable disease surveillance and rapid nutrition surveys are critical tools used to detect potential epidemics and famine¹¹. Unfortunately, a number of factors contribute to data quality issues during a humanitarian crisis including the inability to directly supervise personnel in the field, lack of resources on the ground, clinics or headquarters being overwhelmed, inadequately trained personnel in the field, safety concerns in the field, and poor communication with headquarters. In many cases, the data analyst is withdrawn from the on-site data collection process and is asked to analyze data without knowledge of the quality of the reporting system. The analyst needs a method of assessing data quality based on the database alone to inform effective intervention decisions. The goal of this

study is to develop methods to allow the data analyst to assess the data quality from surveillance tallies submitted by reporting facilities.

Quality assessment checks using the Emergency Nutrition Assessment (ENA) software have already been developed for emergency nutrition surveys¹². The checks look for age and sex distribution, normality, kurtosis and skew of the weight-for-height Z-score distribution, distribution of malnutrition cases across clusters based on the Poisson distribution, and digit preference scores for children's heights and weights. These checks were successfully implemented in nutrition and mortality surveys from 12 districts during the 2011 Somalia famine¹³. Implementing a comparable system for emergency surveillance systems would prove beneficial.

Since 2011, the CDC has assisted WHO Somalia in analyzing the Somalia disease surveillance system¹⁰. Some of the data were difficult to analyze and a number of surveys and clinics had to be dropped from the analysis due to severe issues in data quality. From 2011-2013, health facilities reporting to the WHO Somalia CSR system were evaluated using four quality domains: completeness, internal consistency, duplication, and outliers. In 2012, an ACCESS based data entry system was introduced to replace the old Microsoft Excel system. This led to a reduction in errors and an overall improvement in the quality of reported data.

Using the work on the Somalia surveillance system, we have refined the methods of assessing quality and developed an index of quality for each facility. We evaluated the current methods and added additional methods including those borrowed from ENA. Developing a standard index of quality will be a useful tool for the analysis of surveillance data from future humanitarian crises and as a feedback mechanism to improve data quality.

2. Methods

The data originated from the WHO's Somalia CSR. Selected hospitals, maternal and child healthcare facilities (MCH), and other healthcare facilities contributed to weekly reports. Case information was entered into standardized registries located at healthcare facilities. Providers at healthcare facilities then entered summary surveillance data into weekly tally sheets. The weekly tally sheets were picked up by regional health workers and delivered to the zonal level for entry into surveillance databases. After consent from the WHO Country Representative, the yearly data were provided to the Emergency Response and Recovery Branch (ERRB) of the Centers for Disease Control and Prevention (CDC). Post-hoc analysis for data quality issues was performed on the aggregated electronic databases.

Midway through 2012, the CSR system was improved with additional quality assurance checks. Data for the first part of 2012 were entered in Microsoft Excel (Central Zone weeks 1-13, Southern Zone weeks 1-13, Puntland weeks 1-24, Somaliland weeks 1-23), while data for the latter part of the year were available in a Microsoft Access file that was developed by ERRB as an Epi Info™ 7 database (CDC, Atlanta, GA). The two files for each zone were merged for cleaning and analysis using SAS (version 9.3). The 2013 data were available in Microsoft Access in its entirety. Additionally, a number of facilities were dropped after 2012 data collection because of poor reporting. In total, 234 facilities were included in the 2012 database. Subsequently, WHO Somalia dropped 36 facilities, leaving 198 facilities for the 2013 analysis.

Somalia health facilities were allocated to one of four zones: Southern, Central, Somaliland or Puntland. The data cover all 52 weeks of the year and consist of 9 conditions: suspected shigellosis, acute flaccid paralysis (an indicator of polio), suspected measles, confirmed malaria, suspected cholera, suspected

pertussis, neonatal tetanus, suspected diphtheria, and all other consultations. Data were further broken down such that the number of male and female cases for each health event and the cases of priority health events among individuals under 5 years of age and 5 years of age and older were reported. Additionally, a variable for total consults was reported. Visits for any of the above 9 conditions were counted under this variable.

Assessing Data Quality: Within each facility, the data were assessed using a number of different quality attributes. Each attribute was scored between 0 and 1, with low scores being better. The quality attributes were summed and a final overall data quality score between 0 and 10 was given. The quality attributes are listed below and exact calculations are explained in Sections 2.1-2.4.

- 1) Completeness of the data
 - a. Proportion of missing weeks
 - b. Proportion of weeks with all zeros
- 2) Internal Consistency
 - a. Sex ratio tests on disease case counts
 - b. Proportion of reporting mistakes (not abiding by case definitions)
 - c. Proportion of weeks where disease counts plus all other consultations do not sum to total consults
 - d. Digit preference tests on disease case counts
- 3) Duplication
 - a. Proportion of duplicate records for facility and week
 - b. Proportion of duplicate counts for all diseases when compared with a previous week
- 4) Outliers
 - a. Mean and standard deviation method
 - b. Index of dispersion method

2.1 Completeness of the Data

Data were assessed based on completeness of records within each facility. As criteria for completeness, the proportion of weeks in the year left unreported (“missing weeks”) was calculated.

$$MW_i = \frac{\text{number of missing weeks}}{52}$$

In addition, weeks in which cases counts for all diseases, including total consults, were reported as 0 were identified. These weeks were considered to have missing information. We used the proportion of all such weeks (“zero weeks”) as a second measure of completeness.

$$ZW_i = \frac{\text{number of weeks with all zeros}}{\text{total number of reported weeks}}$$

2.2 Internal Consistency

To detect unusual patterns in the data, digit preference tests were conducted on total case counts and all other consultation counts for each facility and chi squared statistics were calculated. For these variables, which had a high number of cases, it was assumed that the last digit of the cases counts would follow a uniform distribution. The digit preference test uses the chi squared statistic to measure the deviation of the actual distribution of case counts from the expected uniform distribution. Since our data is categorical and we are comparing to an expected categorical distribution, the chi squared

statistic is appropriate. The degrees of freedom of each test will equal $n-1$, where n is the number of final digits that appeared in that facility's case counts. Over the 52 weeks, most facilities will have 10 final digits (0, 1, ... ,9). In this case the degrees of freedom of the test was 9. Some tests had fewer degrees of freedom because that facility did not have all 10 final digits represented. In our metric, the digit preference attribute calculates the proportion of tested variables that were found to have significant scores on the digit preference test. We tested at the .00625 (.05/8) level based on Bonferroni's adjustment for multiple comparisons.

$$DP_i = \text{proportion of variables with significant } \chi^2 = \sum \left(\frac{(\text{observed} - \text{expected})^2}{\text{expected}} \right) \text{ value}$$

As an additional method, we calculated the overall p-value for the digit preference score for each facility. Based on the properties of the chi squared distribution, independent χ^2 values will sum to another χ^2 variable, and its degrees of freedom will also sum¹⁴. This can be proved using moment generating functions (see Taboga, M. 2010). Let X_1 be a chi-square random variable with v_1 degrees of freedom, X_2 be a chi-square random variable with v_2 degrees of freedom and so on. The moment generating function of X_i is

$$M_{x_i}(t) = (1 - 2t)^{-v_i/2}$$

And we can define X as the sum of our independent chi square variables.

$$X = \sum_{i=1}^k X_i$$

The moment generating function of the sum of mutually independent random variables is the product of their moment generating functions.

$$\begin{aligned} M_X(t) &= \prod_{i=1}^k M_{x_i}(t) \\ &= \prod_{i=1}^k (1 - 2t)^{-v_i/2} \\ &= (1 - 2t)^{-\sum_{i=1}^k v_i/2} \\ &= (1 - 2t)^{-v/2} \end{aligned}$$

$$\text{where } v = \sum_{i=1}^k v_i$$

Since the moment generating function of X is the moment generating function of a chi-square variable with v degrees of freedom, X is a chi square variable with v degrees of freedom¹⁵

■

We also know that if $X \sim \chi^2(v)$ then $E(X) = v$. It follows that $E(X/v)$ is then equal to 1. We summed the χ^2 and v values for each of our digit preference tests $(\chi^2(v_1) + \dots + \chi^2(v_8)) / (v_1 + \dots + v_8)$ and found an overall χ^2/v value with an expected value of 1. Using that statistic we found overall digit preference p-values for each facility.

As another measure of internal consistency, we tested each facility for unusual patterns in the sex ratio of their weekly case counts. To do this, each facility was assumed to have some gender distribution (as calculated by the total gender distribution at that facility over all weeks). We then conducted chi-square

goodness of fit tests to test for weekly gender distributions that varied significantly from the overall trend. In this case, the degrees of freedom for each test equaled 1 (degrees of freedom = n-1 where n=2 for our 2 categories of gender). The proportion of individual weeks which showed significant deviation were recorded, again correcting for multiple comparisons. We also calculated overall sex ratio p-values at each facility.

$$SR_i = \text{proportion of weeks with significant } \chi^2 = \sum \left(\frac{(\text{observed} - \text{expected})^2}{\text{expected}} \right) \text{ value}$$

For our third measure of internal consistency, we summed the total counts for each individual disease over each week in any given facility. We then compared that number to the count the facility recorded as the total consults variable. Since each reported condition should have also been reported under the total consults variable, we would expect the summed counts to equal the total consults count. Any discrepancies were marked and we calculated the proportion of weeks where the counts were not equal (“non-matching weeks”).

$$NMW_i = \frac{\text{sum of weeks where totals did not match (in 5 over and variable and less than 5 variable)}}{\text{number of total reported weeks between the 2 variables}}$$

Finally, we checked for instances in which facilities reported cases of neonatal tetanus in the 5 years and over age category. By the case definition, these cases should not have been reported and represent facility data inconsistencies.

$$RM_i = \frac{\text{weeks with reported values for tetanus over 5}}{\text{number of total reported weeks}}$$

2.3 Duplication

Duplication in the data came in two forms. First, there were instances where two records were reported for same week, but the records contained conflicting case counts. These weeks were classified as “double reporting weeks”. The proportion of these weeks was the first measure of duplication.

$$DRW_i = \frac{\text{number of weeks with identical facility and week variable}}{52}$$

Second, there were instances where different weeks in a facility had the exact same case counts for all diseases. These were flagged as potentially erroneous records. Weeks in which all disease reports of four or more cases were identical to that of a previous week were classified “duplicate case count weeks”. Duplicate case count weeks were identified by examining all reporting weeks within each facility. When multiple weeks were found to have identical case counts, the first week was considered valid, but any subsequent weeks were considered to be potentially erroneous data. The proportion of duplicate weeks was recorded for each facility.

$$DCCW_i = \frac{\text{number of weeks with identical case counts in a previous week}}{\text{number of total reported weeks}}$$

2.4 Outliers

Case counts for these health events were considered outliers if case counts within a given week were greater than 3 standard deviations away from the mean of all case reports of respective health events for that facility. As an additional requirement, these weeks had to have a case count greater than 10 and both the previous week or the next week had to have less than 40% of the cases of the current week. The proportion of weeks with detected outliers was calculated.

$$OW_i = \frac{\text{number of variables with detected outliers}}{\text{total number of variables tested}}$$

Additionally, we calculated the index of dispersion based on the distribution of case counts of each of the 8 disease conditions at each facility. For cases of a disease in a period of time (0, t], The index of dispersion is calculated as $V(t)/M(t)$, where $V(t)$ is the variance of the case counts and $M(t)$ is the mean of the case counts¹⁶. In the Poisson distribution, the expected value and the variance are both equal to λ ¹⁷. Thus the expected value of the index of dispersion = $\frac{\lambda}{\lambda} = 1$. In effect, the index of dispersion measures the ratio of $V(t)$ to its value from a Poisson process. A negative binomial distribution (over-dispersed Poisson) can be used as an alternative to the Poisson distribution in cases where the sample variance exceeds the sample mean¹⁸. In this case the index of dispersion would be greater than 1. Either way, unusually high values for the index of dispersion are warning signs for potential outliers and other reporting problems. In our metric, the index of dispersion attribute calculates the proportion of tested variables that had an unusually high index of dispersion (index of dispersion > 5) in each facility.

$$IOD_i = \frac{\text{number of variables with index of dispersion over 5}}{\text{total number of variables tested}}$$

2.5 Testing Differences Between Zones

We ran a one-way analysis of variance (ANOVA) to test for differences in data quality scores between the four zones.

3. Results

3.1 Southern Zone

The overall data quality scores for each of our 198 facilities, as well as the facility scores on each quality attribute, were calculated and summarized. We will present the results by zone, beginning with results from Southern Zone, which are shown below (Table 1). Of the 37 facilities in Southern Zone, data quality scores ranged from 0 to 1.070. The scores were found to be skewed to the right, with a mean score of 0.273 and a median score of 0.188 (sd = 0.293). Of all the zones, Southern Zone had the most serious problems with missing weeks. Southern Zone had 229 missing weeks, compared to 131 in Central Zone, 63 in Puntland and 58 in Somaliland.

Table 1. Data Quality Scores for Southern Facilities (Facilities 1-37)

Facility	1. Completeness		2. Internal Consistency				3. Duplication		4. Outliers		Overall Facility Score
	Missing Weeks	Weeks with all zeros	Sex ratio of the diseases	Reporting mistakes (tetanus over 5)	Non-matching weeks	Digit preference	Duplicate records for facility and week	Duplicate case counts week	Mean and standard deviation method	Index of Dispersion	
SZLJ08	0.808	0	0.200	0	0	0	0	0	0	0.063	1.070
SZLJ01	0	0	0.077	0	0.019	0.250	0	0	0	0.125	0.471
SZBY07	0	0	0	0	0.019	0	0	0	0	0	0.019
SZLJ03	0.904	0	0	0	0	0	0	0	0	0	0.904
SZBY02	0	0	0.308	0	0.019	0	0	0.019	0.019	0	0.365
SZBY01	0	0	0	0	0	0	0	0	0	0	0
SZGE03	0	0	0.038	0.019	0.029	0	0	0	0.038	0.063	0.188
SZGE02	0	0	0.308	0	0.058	0.125	0	0	0	0	0.490
SZLJ09	0.904	0	0	0	0	0	0	0	0	0	0.904
SZGE08	0.154	0	0.023	0	0	0	0	0	0	0	0.177
SZBY06	0	0	0.115	0	0	0	0	0	0	0	0.115
SZLJ02	0	0	0.038	0	0.019	0.125	0	0	0.019	0	0.202
SZGE04	0.019	0	0.137	0	0.020	0	0	0.020	0	0	0.196
SZGE11	0	0	0.192	0	0	0	0	0.058	0	0	0.250
SZGE06	0	0	0.058	0	0.029	0	0	0	0	0	0.087
SZGE07	0.019	0	0	0	0	0	0	0	0	0	0.019
SZLJ10	0	0	0.115	0	0.010	0.125	0	0	0	0	0.250
SZGE10	0.019	0	0.020	0	0	0.125	0	0.039	0	0	0.203
SZGE05	0	0	0.058	0	0	0	0	0.019	0	0	0.077
SZGE01	0	0	0.019	0	0	0.125	0	0	0	0	0.144
SZLJ05	0	0	0.019	0	0.010	0	0	0	0	0	0.029
SZBK01	0	0	0.019	0	0.019	0	0	0	0.019	0.063	0.120
SZMJ01	0	0	0	0	0.010	0	0	0	0	0	0.010
SZLJ06	0	0	0	0	0.010	0.250	0	0.019	0	0.063	0.341
SZLJ11	0	0	0.038	0	0.019	0	0	0.019	0.019	0	0.096
SZLJ04	0.904	0	0	0	0	0	0	0	0	0	0.904
SZBY04	0	0	0.019	0	0	0	0	0.019	0	0	0.038
SZGE12	0	0	0.192	0	0.010	0	0	0	0	0	0.202
SZBY05	0	0	0.096	0	0.038	0	0	0	0	0.125	0.260
SZLJ07	0	0	0.038	0	0	0	0	0	0	0.063	0.101
SZLJ12	0.019	0	0.333	0	0.020	0	0	0.039	0	0	0.411
SZLJ13	0.654	0	0.111	0	0.056	0	0	0	0	0	0.821
SZBY08	0	0	0.019	0	0	0	0	0	0	0	0.019
SZMJ04	0	0	0.038	0	0.019	0	0	0.038	0	0	0.096
SZBK02	0	0	0	0	0	0	0	0	0	0	0
SZBK03	0	0	0.115	0	0	0	0	0	0	0	0.115
SZMJ05	0	0	0.135	0	0.019	0.250	0	0.019	0	0	0.423

3.2 Central Zone

In Central Zone, data quality scores from 61 facilities ranged from 0 to 1.115. The mean score was 0.252 and the median score was 0.164 (sd=.252). Full score reports are shown below (Tables 2-3). Central Zone had the worst average score on the outlier attributes. 13 outliers were detected, compared to 6 in Southern Zone, 3 in Somaliland and 2 in Puntland. 2.77% of the variables tested had index of dispersion values greater than 5 compared to 1.52% in Southern Zone, 1.22% in Puntland and 0.35% in Somaliland. Additionally, Central Zone had 3 Facilities which did not abide by the case definition for tetanus, compared to 1 facility in Southern Zone, 1 facility in Somaliland and no facilities in Puntland.

Table 2. Data Quality Scores for Central Facilities (Facilities 1-30)

Facility	1. Completeness		2. Internal Consistency				3. Duplication		4. Outliers		Overall Facility Score
	Missing Weeks	Weeks with all zeros	Sex ratio of the diseases	Reporting mistakes (tetanus over 5)	Non-matching weeks	Digit preference	Duplicate records for facility and week	Duplicate case counts week	Mean and standard deviation method	Index of Dispersion	
SZBN11	0	0	0.096	0	0	0.250	0	0	0	0	0.346
SZBN01	0	0	0.538	0	0	0	0	0	0	0	0.538
SZGA01	0	0	0.019	0	0.019	0.125	0	0	0	0	0.163
SZGA02	0	0	0	0.019	0.010	0	0	0	0	0	0.029
SZBN17	0	0	0.135	0	0.019	0.250	0	0	0.019	0.063	0.486
SZLS17	0.500	0	0	0	0	0	0	0.038	0.038	0	0.577
SZLS09	0	0	0.019	0	0.019	0	0	0.019	0.019	0.125	0.202
SZLS14	0	0	0	0	0.019	0	0	0.019	0	0	0.038
SZMS01	0	0	0.173	0	0.019	0.125	0	0	0	0.063	0.380
SZBN18	0	0	0.173	0	0.029	0	0	0	0.038	0.375	0.615
SZLS12	0	0	0.038	0	0.010	0	0	0.019	0	0	0.067
SZLS13	0	0	0.135	0	0	0.125	0	0.019	0	0	0.279
SZLS10	0	0	0.058	0	0.019	0	0	0.019	0	0.063	0.159
SZBN02	0	0	0	0	0	0.125	0	0	0	0	0.125
SZHA01	0	0	0.019	0	0.019	0	0	0	0	0	0.038
SZLS11	0	0	0	0	0	0.125	0	0.019	0	0	0.144
SZBN23	0	0	0	0	0	0.125	0	0.019	0	0	0.144
SZBN03	0	0	0.077	0	0	0.250	0	0	0	0	0.327
SZGA03	0	0	0.019	0	0	0	0	0	0	0	0.019
SZGA04	0	0	0.058	0	0.010	0.250	0	0	0	0	0.317
SZGA05	0	0	0.135	0	0	0	0	0.019	0	0	0.154
SZGA06	0	0	0.192	0	0	0	0	0.038	0	0	0.231
SZLS01	0	0	0.250	0	0	0.625	0	0	0.019	0	0.894
SZGA07	0	0	0.096	0	0.019	0	0	0	0.038	0.125	0.279
SZLS03	0	0	0.115	0	0.019	0	0	0.038	0	0.125	0.298
SZLS18	0	0	0	0	0.010	0	0	0.019	0	0.188	0.216
SZBN05	0	0	0	0	0	0	0	0	0	0	0
SZBN09	0.019	0	0	0	0.010	0	0	0	0	0	0.029
SZBN10	0.019	0	0.020	0	0	0.125	0	0	0	0	0.164
SZBN04	0	0	0	0	0	0	0	0.019	0	0	0.019

Table 3. Data Quality Scores for Central Facilities (Facilities 31-61)

Facility	1. Completeness		2. Internal Consistency				3. Duplication		4. Outliers		Overall Facility Score
	Missing Weeks	Weeks with all zeros	Sex ratio of the diseases	Reporting mistakes (tetanus over 5)	Non-matching weeks	Digit preference	Duplicate records for facility and week	Duplicate case counts week	Mean and standard deviation method	Index of Dispersion	
SZBN12	0	0	0	0	0	0	0	0	0	0	0
SZLS15	0	0	0.038	0	0	0.375	0	0.019	0	0	0.433
SZHA02	0	0	0.192	0	0.010	0	0	0.019	0	0.063	0.284
SZLS19	0	0	0.135	0	0	0.125	0	0.019	0.019	0.063	0.361
SZMS05	0	0	0.096	0	0	0	0	0	0	0	0.096
SZLS20	0.038	0	0	0	0	0	0	0.020	0	0	0.058
SZBN16	0	0	0	0.019	0	0	0	0	0	0.063	0.082
SZLS16	0	0	0.135	0	0	0	0	0.019	0	0	0.154
SZBN24	0.981	0	0	0	0	0	0	0	0	0	0.981
SZLS06	0	0	0.135	0	0.019	0	0	0.038	0	0	0.192
SZBN28	0	0	0.019	0	0	0	0	0	0	0	0.019
SZBN20	0	0	0.058	0.019	0.010	0.125	0	0	0	0.063	0.274
SZBN21	0.038	0	0.140	0	0	0.250	0	0.020	0.020	0.125	0.593
SZLS21	0.154	0	0.068	0	0.011	0	0	0	0	0.063	0.296
SZBN29	0	0	0.019	0	0	0	0	0	0	0	0.019
SZMS06	0.731	0	0.071	0	0	0	0	0.071	0	0	0.874
SZLS22	0	0	0	0	0.019	0	0	0.019	0	0	0.038
SZLS26	0	0	0.019	0	0	0.250	0	0.077	0	0	0.346
SZLS27	0	0	0.019	0	0	0	0	0.019	0	0	0.038
SZBN22	0	0	0.115	0	0	1.000	0	0	0	0	1.115
SZBN14	0	0	0.192	0	0.019	0	0	0	0.019	0.125	0.356
SZLS28	0	0	0	0	0	0	0	0.019	0	0	0.019
SZLS23	0	0	0.038	0	0.029	0	0	0	0	0	0.067
SZLS24	0.038	0	0.240	0	0	0	0	0.020	0	0	0.298
SZLS08	0	0	0.385	0	0	0	0	0.019	0	0	0.404
SZBN26	0	0	0.077	0	0	0	0	0	0	0	0.077
SZLS30	0	0	0.038	0	0	0	0	0.038	0.038	0	0.115
SZLS29	0	0	0	0	0	0	0	0.038	0	0	0.038
SZMS08	0	0	0	0	0	0	0	0.019	0	0	0.019
SZBN30	0	0	0.135	0	0	0.250	0	0	0	0	0.385
SZBN13	0	0	0.058	0	0	0	0	0	0	0	0.058

3.3 Puntland Zone

In Puntland, data quality scores from 46 facilities ranged from 0 to 1.304. The mean score was 0.223 and the median score was 0.115 (sd=.223). Full score reports are shown below (Tables 4-5).

Puntland facilities had the worst average score for the digit preference attribute. Over the course of analysis, we found that 41.30% of Puntland facilities had potential digit preference problems compared to 36.07% of facilities in Central Zone, 27.00% of facilities in Southern Zone, 33.33% of facilities in Somaliland.

Table 4. Data Quality Scores for Puntland Facilities (Facilities 1-25)

Facility	1. Completeness		2. Internal Consistency				3. Duplication		4. Outliers		Overall Facility Score
	Missing Weeks	Weeks with all zeros	Sex ratio of the diseases	Reporting mistakes (tetanus over 5)	Non-matching weeks	Digit preference	Duplicate records for facility and week	Duplicate case counts week	Mean and standard deviation method	Index of Dispersion	
PLMU04	0.019	0	0.020	0	0	0.500	0	0.039	0	0	0.578
PLMU09	0	0	0.115	0	0	0	0	0	0	0	0.115
PLKA01	0.019	0	0.019	0	0	0	0.038	0	0	0	0.077
PLSA01	0	0	0.019	0	0	0	0	0.038	0	0	0.058
PLSA02	0.019	0.019	0	0	0	0.250	0.019	0.096	0	0	0.404
PLSL01	0	0	0	0	0	0.250	0	0	0	0	0.250
PLBA03	0	0.058	0.154	0	0	0.125	0	0.019	0	0.125	0.481
PLBA01	0.019	0	0.275	0	0.020	0.125	0	0.020	0	0.125	0.583
PLBA02	0	0	0.170	0	0.028	0	0.019	0	0	0.063	0.280
PLCA01	0	0	0	0	0	0.250	0	0	0	0	0.250
PLNG01	0	0	0.019	0	0	0	0	0	0	0	0.019
PLBA04	0	0	0.038	0	0	0	0	0	0	0	0.038
PLNG04	0	0.019	0.057	0	0	0.500	0.019	0.057	0	0	0.651
PLNG03	0	0	0	0	0	0.125	0.038	0	0	0.063	0.226
PLSA04	0	0	0	0	0.019	0	0	0.038	0	0	0.058
PLKA02	0	0	0.019	0	0	0.250	0	0.038	0	0	0.308
PLMU05	0	0	0.308	0	0	0	0	0.019	0	0	0.327
PLMU03	0	0	0.327	0	0	0.625	0	0	0	0	0.952
PLMU06	0.096	0	0.234	0	0	0.125	0	0	0.021	0.063	0.539
PLNG08	0.019	0	0.020	0	0.029	0	0	0	0	0	0.068
PLNG09	0	0.019	0.075	0	0	0	0.019	0	0	0	0.114
PLNG05	0	0	0.038	0	0	0.500	0	0.019	0	0	0.558
PLSA03	0	0	0.019	0	0	0	0	0	0	0	0.019
PLKA03	0.019	0	0.039	0	0	0	0	0.020	0	0	0.078
PLMU07	0	0	0.135	0	0	0	0	0	0	0	0.135

Table 5. Data Quality Scores for Puntland Facilities (Facilities 26-46)

Facility	1. Completeness		2. Internal Consistency				3. Duplication		4. Outliers		Overall Facility Score
	Missing Weeks	Weeks with all zeros	Sex ratio of the diseases	Reporting mistakes (tetanus over 5)	Non-matching weeks	Digit preference	Duplicate records for facility and week	Duplicate case counts week	Mean and standard deviation method	Index of Dispersion	
PLSL02	0	0	0	0	0	0	0	0	0	0	0
PLBA05	0.038	0	0.019	0	0.019	0	0.038	0	0	0	0.115
PLNG02	0	0	0	0	0.019	0	0	0	0	0	0.019
PLMU10	0	0	0.135	0	0	0	0	0.019	0	0	0.154
PLNG10	0.038	0	0	0	0	0	0	0	0	0	0.038
PLSL03	0	0	0	0	0	0	0	0	0	0	0
PLSL04	0	0	0	0	0	0	0	0.019	0	0	0.019
PLSL05	0	0	0	0	0	0.250	0	0	0	0	0.250
PLMU02	0.904	0	0.400	0	0	0	0	0	0	0	1.304
PLNG11	0	0	0.058	0	0.010	0	0	0	0	0	0.067
PLKA04	0	0.019	0.038	0	0	0	0	0	0	0	0.058
PLKA05	0	0	0.058	0	0.029	0	0	0	0	0	0.087
PLNG06	0	0	0	0	0.010	0	0	0	0	0	0.010
PLKA06	0	0	0.019	0	0	0	0	0	0.019	0.125	0.163
PLKA07	0	0	0.019	0	0	0	0	0.019	0	0	0.038
PLNG12	0	0.019	0	0	0	0.250	0	0.019	0	0	0.288
PLSL06	0	0	0	0	0	0	0	0.019	0	0	0.019
PLBA06	0	0	0	0	0	0	0	0.019	0	0	0.019
PLKA08	0	0	0.019	0	0	0.250	0	0	0	0	0.269
PLNG07	0.019	0.019	0.038	0	0	0	0.038	0	0	0	0.114
PLSA05	0	0.058	0	0	0	0	0	0.019	0	0	0.077

3.4 Somaliland Zone

In Somaliland, data quality scores from 54 facilities ranged from 0.192 to 1.418. The mean score was 0.220 and the median score was 0.092 (sd=.277). Full score reports are shown below (Tables 6-7). Somaliland showed the most serious problems with duplicate case count weeks. A total of 115 weeks with duplicated case counts were found. In comparison, Central Zone had 38 duplicate case count weeks, Southern Zone had 26 and Puntland had 25. Somaliland also had the highest number of 'non-matching weeks' (weeks where the summed case counts for each individual disease condition did not add up to the total consultations count). Somaliland had 63 such weeks, compared to 31 in Southern Zone, 29 in Central Zone and 13 in Puntland.

Table 6. Data Quality Scores for Somaliland Facilities (Facilities 1-26)

Facility	1. Completeness		2. Internal Consistency				3. Duplication		4. Outliers		Overall Facility Score
	Missing Weeks	Weeks with all zeros	Sex ratio of the diseases	Reporting mistakes (tetanus over 5)	Non-matching weeks	Digit preference	Duplicate records for facility and week	Duplicate case counts week	Mean and standard deviation method	Index of Dispersion	
SLMA06	0.019	0.020	0	0	0.010	0.250	0	0	0	0	0.299
SLTO01	0.019	0	0.167	0	0.028	0.375	0.058	0.167	0	0	0.813
SLMA07	0.019	0	0	0	0.039	0	0	0	0	0	0.058
SLSN01	0	0	0.058	0	0.010	0	0	0.038	0	0	0.106
SLMA01	0.019	0	0	0	0.020	0	0	0.020	0	0	0.058
SLMA02	0.038	0	0.060	0	0	0	0	0.020	0	0	0.118
SLSN04	0	0	0.058	0	0	0	0	0.038	0	0	0.096
SLSH01	0	0	0.038	0	0	0	0	0	0	0	0.038
SLSH04	0	0	0	0	0.029	0	0	0	0	0	0.029
SLAW02	0	0	0.019	0	0.019	0	0.019	0	0	0	0.057
SLAW03	0	0	0	0	0.029	0.125	0	0	0	0	0.154
SLTO02	0.038	0	0.176	0	0.010	0.250	0.019	0.118	0	0	0.612
SLTO03	0.038	0	0.160	0	0	0.250	0	0.200	0	0	0.648
SLTO04	0.019	0	0.176	0	0.029	0.250	0	0.078	0	0	0.554
SLAW04	0	0	0.019	0	0	0	0	0	0	0	0.019
SLMA08	0.019	0	0	0	0.048	0	0.019	0	0	0	0.087
SLSO03	0.019	0	0.039	0	0.029	0	0	0.020	0	0	0.107
SLMA09	0.058	0	0.120	0	0.010	0.125	0.019	0.020	0	0	0.352
SLSN02	0	0	0	0	0.019	0	0	0.058	0	0	0.077
SLAW01	0.019	0	0.019	0	0.019	0	0.019	0	0	0	0.077
SLTO05	0.038	0	0.180	0	0	0	0	0.180	0	0	0.398
SLTO06	0.038	0	0.078	0	0	0.125	0.019	0.176	0	0	0.438
SLSN03	0.019	0	0.020	0	0	0	0	0.039	0.020	0.063	0.160
SLSN05	0.019	0.170	0.019	0	0.028	0.250	0.038	0	0	0	0.525
SLSO04	0.019	0	0	0	0.010	0	0	0.020	0	0	0.049
SLTO07	0.038	0	0.160	0	0.020	1.000	0	0.200	0	0	1.418

Table 7. Data Quality Scores for Somaliland Facilities (Facilities 27-54)

Facility	1. Completeness		2. Internal Consistency				3. Duplication		4. Outliers		Overall
	Missing Weeks	Weeks with all zeros	Sex ratio of the diseases	Reporting mistakes (tetanus over 5)	Non-matching weeks	Digit preference	Duplicate records for facility and week	Duplicate case counts week	Mean and standard deviation method	Index of Dispersion	Facility Score
SLMA03	0.019	0.078	0.020	0	0.069	0	0	0	0	0	0.186
SLMA04	0.038	0	0	0	0	0	0.019	0.039	0	0	0.097
SLAW07	0.038	0	0	0	0	0	0.019	0	0	0	0.058
SLAW05	0.038	0	0	0.020	0.020	0	0.019	0	0.039	0.125	0.261
SLMA10	0.019	0	0	0	0	0	0.019	0	0	0	0.038
SLMA11	0.019	0	0.020	0	0.039	0	0	0	0	0	0.078
SLMA12	0.038	0	0	0	0.010	0	0	0	0	0	0.048
SLSO01	0.058	0	0.020	0	0.010	0.125	0	0.020	0	0	0.234
SLMA13	0.019	0	0	0	0.010	0	0.019	0	0	0	0.048
SLAW08	0.019	0	0.019	0	0.010	0	0.019	0	0	0	0.067
SLSH02	0	0	0	0	0	0	0.019	0	0	0	0.019
SLSO02	0.019	0	0	0	0.010	0	0	0.039	0	0	0.068
SLTO08	0.019	0	0.216	0	0.010	0.125	0	0.157	0	0	0.527
SLTO09	0.019	0	0.098	0	0	0.250	0	0.235	0	0	0.603
SLSN06	0	0	0.038	0	0.019	0	0	0.038	0	0	0.096
SLSO05	0.038	0	0	0	0.040	0.125	0	0.020	0	0	0.223
SLSO06	0.019	0	0.019	0	0.019	0	0.038	0.019	0	0	0.114
SLMA14	0.019	0	0.039	0	0	0	0	0	0	0	0.058
SLMA15	0.019	0	0	0	0.020	0	0	0	0	0	0.039
SLTO10	0.019	0	0.275	0	0	0.500	0	0.137	0	0	0.931
SLSH03	0.019	0	0.039	0	0	0	0	0	0	0	0.058
SLMA16	0.019	0	0.020	0	0.010	0	0	0.020	0	0	0.068
SLMA17	0.019	0	0.020	0	0	0	0	0	0	0	0.039
SLAW06	0	0	0.019	0	0	0	0	0.019	0	0	0.038
SLSH05	0.019	0	0	0	0	0	0	0.020	0	0	0.039
SLSO08	0.019	0	0	0	0.020	0	0	0.020	0	0	0.058
SLMA05	0.019	0	0.020	0	0.049	0	0	0	0	0	0.088
SLSO07	0.019	0.020	0	0	0	0.250	0	0.078	0	0	0.367

3.5 Overall Comparisons

Over all facilities the data quality scores ranged from 0 to 1.418 (mean=.241, median= 0.130, sd =.270), with higher scores indicating more severe data quality issues. On average, facilities in the Southern Zone had the worst data quality scores (mean=.273) while facilities in Somaliland had the best scores (mean=.220). Basic summary statistics for data quality scores are shown below (Table 8). Additionally, distributions of scores in each of Somalia’s four zones are shown in the Figures and Tables section (Figures 1-5). We found no significant differences between the data quality scores in different zones (F=0.38, p=0.77).

Table 8. Mean, standard deviation, minimum and maximum scores by zone

Zone	Number of Facilites	Mean	Median	Standard Deviation	Minimum	Maximum
Central	61	0.252	0.164	0.253	0.000	1.115
Puntland	46	0.223	0.115	0.269	0.000	1.304
Somalila	54	0.220	0.092	0.277	0.019	1.418
Southern	37	0.273	0.188	0.293	0.000	1.070
Total	198	0.241	0.130	0.270	0.000	1.418

In an effort to characterize data quality based on our scores, we have proposed a classification system for surveillance data from humanitarian emergencies (Table 9). This table is merely a proposed classification and should be reevaluated and refined after analyzing additional datasets. Comparing data quality from the CSR surveillance system to data from a ‘gold standard’ system is recommended.

Table 9. Proposed Classification of Data Quality Scores

Data Quality Score	Classification	Description
0-0.2	Good quality	Few data quality issues found
0.2-0.4	Moderately good quality	Some quality issues found
0.4-0.6	Moderately poor quality	Some reasonably serious quality issues found
Above 0.6	Poor quality	Serious data quality issues found

Of all four zones, Southern Zone had the highest percentage of facilities that were classified as ‘poor quality’ (13.51%), followed by Somaliland (11.11%), Puntland (8.70%) and Central Zone (5.56%). 21.62% of facilities in the Southern Zone were considered either ‘poor quality’ or ‘moderately poor quality’, compared to 19.57% of Puntland facilities, 18.52% of Somaliland facilities, and 14.76% of Central Zone facilities.

Table 10. Data quality classification by zone (with percentage of classification per facility)

Data Quality	Zone				Total
	Central	Puntland	Somalila	Southern	
Good Quality	34 56%	28 61%	37 69%	20 54%	119
Moderately Good Quality	18 30%	9 20%	7 13%	9 24%	43
Moderately Poor Quality	5 8%	5 11%	4 7%	3 8%	17
Poor Quality	4 7%	4 9%	6 11%	5 14%	19
Total	61	46	54	37	198

Full reports for each facility, including breakdowns of specific weeks with data quality issues and digit preference tables are provided in the appendix (Tables 11-31).

4. Discussion

It should be noted that some of the components we examined are only *suggestive* of data quality problems. The purpose of this study was to set up automated checks to quickly judge data quality based on the data itself. The methods used will find unusual patterns in the data, but careful examination of the patterns is required to confirm an unusual pattern is actually attributable to some limitation in the surveillance system. Conversely, a good data quality score does not necessarily guarantee good data. Any method of assessing data quality without knowledge of the entire data collection process will be limited in this way. Data quality assessment should always be an explorative process.

The results of the digit preference tests were among the more interesting findings in our study. A number of facilities had unusual patterns and in some instances, facilities were clearly caught rounding case counts to the nearest 5 or 10. Multiple methods for scoring digit preference were considered and it is highly recommended that digit preference tables for all facilities (Tables 20-24) are examined closely, even for facilities with good scores on that attribute.

Another interesting question we encountered was how to best choose a method for outlier detection. Using only the data itself, there is no definitive way to demonstrate that an extreme observation is an outlier. Observations that appear to be outliers can be very important in surveillance data, signaling that the surveillance system is working correctly to identify an outbreak. Still, with the Somalia data, which was gathered during a humanitarian emergency, it is likely that some outliers were due to issues with the data collection or data entry. We tried to use strict criteria so that the suspected outliers we found were unlikely to be part of a true outbreak and were much more likely to be bad data. A number of changes could be made to the criteria depending on the desired sensitivity of the outlier detection system.

We considered using a time-series approach as an additional way to detect outliers but we ultimately did not use this approach because we felt that there was not enough data to model the seasonal patterns of our diseases. A minimum of 50 observations and several season's worth of data are

recommended to fit an ARIMA model¹⁹. In protracted humanitarian emergencies with larger datasets, a time-series approach could be a valuable addition to the methods used in this paper.

An attribute which should be considered for future use is missing conditions within a week. In our dataset if a week was reported then we had counts for every disease condition within the week. This was a strength of our system (if we assume all disease counts were recorded and reported accurately). In some other datasets there will likely be missing observations within a reported week. This is a data quality issue which should be addressed.

5. Conclusion

We found a wide range of data quality among participating facilities and a number of data quality issues that should be addressed. Interestingly, we did not find significant differences between zones. In the future, we hope that this index can be used to analyze other datasets containing surveillance disease data. With additional practice there can be further refinement of the current methods.

We feel that this index can be a valuable way to evaluate and improve surveillance in humanitarian emergencies. It will allow the data analyst to detect inconsistencies in the data and relay that information to managers overseeing the surveillance system. From there, changes can be made that will ultimately improve the ability of those on the ground to reduce mortality and morbidity.

6. Figures and Tables

Figure 1. Distribution of data quality score by zone

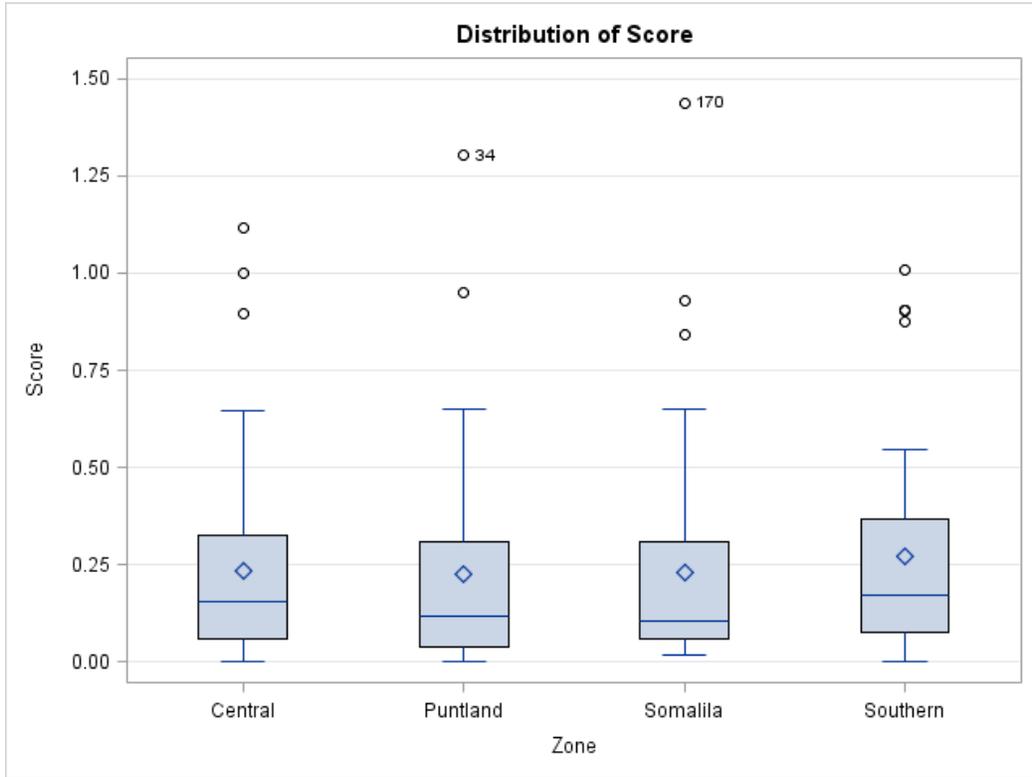


Figure 2. Distribution of data quality scores, Southern Zone facilities

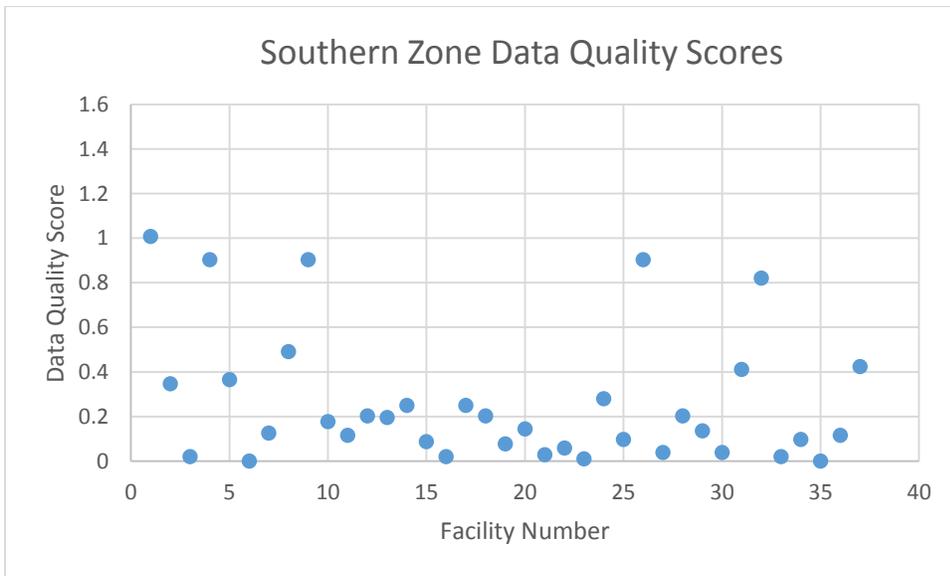


Figure 3. Distribution of data quality scores, Central Zone facilities

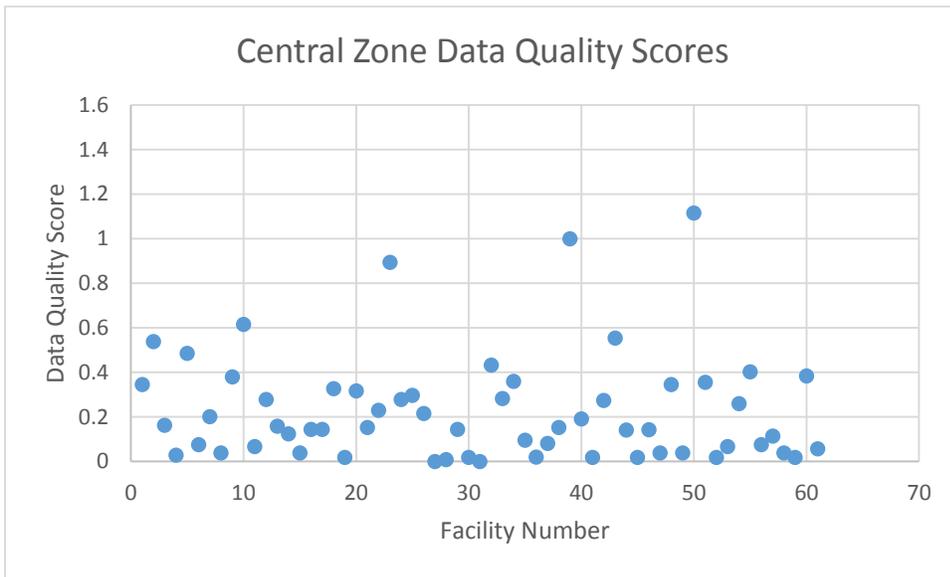


Figure 4. Distribution of data quality scores, Puntland Zone facilities

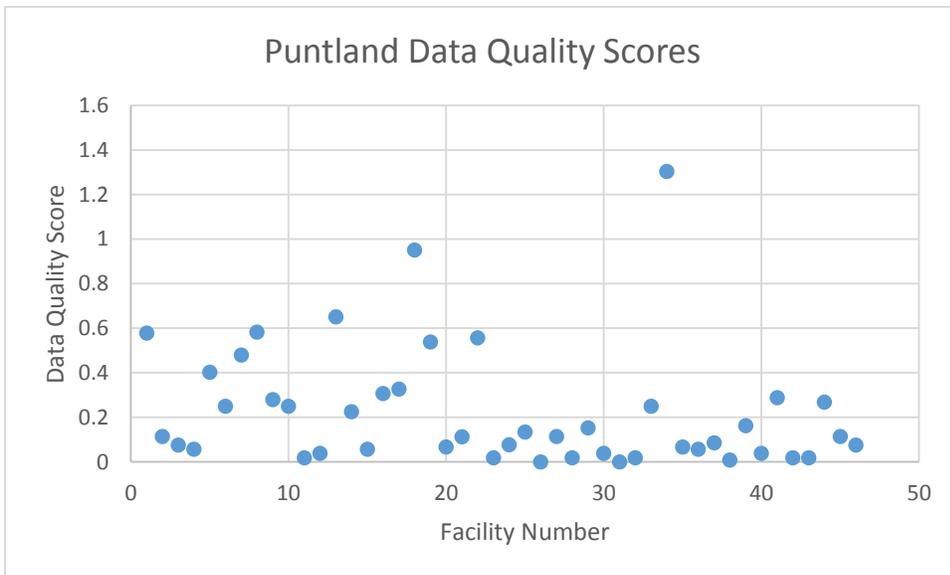


Figure 5. Distribution of data quality scores, Somaliland Zone facilities

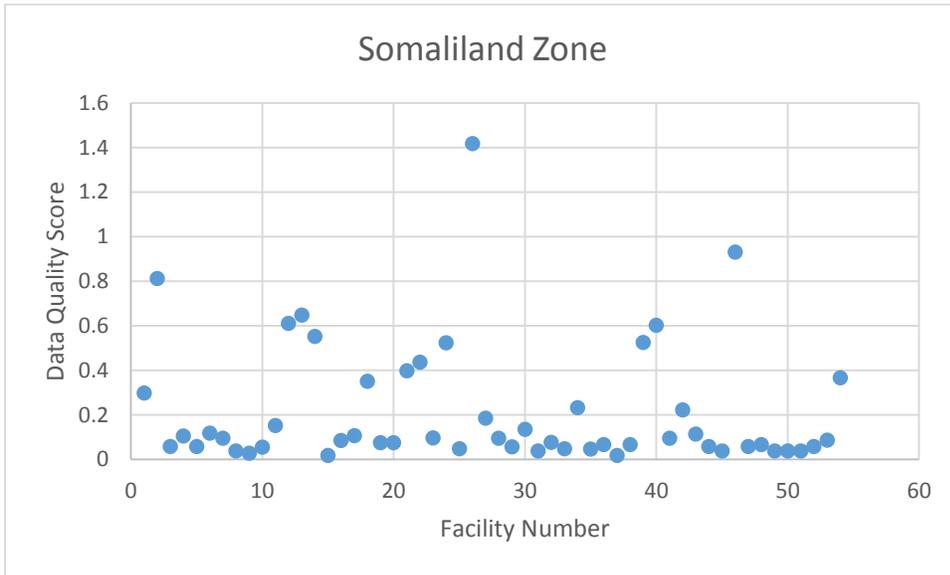


Table 11. Number of reports, number of missing reports, number of duplicate weeks, number of weeks with all zeros, and number of duplicate case count weeks in Southern facilities (facilities 1-37)

Facility	Number of Reports	Total Missing Weeks	% Missing Weeks	Total Number of Weeks with all Zeros	% Weeks with all Zeros	Total Duplicate Weeks	% Duplicate Weeks	Duplicate Case Count Weeks	% Duplicate Case Count Weeks
SZLJ08	10	42	80.77%	0	0.00%	0	0.00%	0	0.00%
SZLJ01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBY07	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLJ03	5	47	90.38%	0	0.00%	0	0.00%	0	0.00%
SZBY02	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBY01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGE03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGE02	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLJ09	5	47	90.38%	0	0.00%	0	0.00%	0	0.00%
SZGE08	44	8	15.38%	0	0.00%	0	0.00%	0	0.00%
SZBY06	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLJ02	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGE04	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
SZGE11	52	0	0.00%	0	0.00%	0	0.00%	3	5.77%
SZGE06	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGE07	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SZLJ10	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGE10	51	1	1.92%	0	0.00%	0	0.00%	2	3.92%
SZGE05	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZGE01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLJ05	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBK01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZMJ01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLJ06	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLJ11	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLJ04	5	47	90.38%	0	0.00%	0	0.00%	0	0.00%
SZBY04	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZGE12	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBY05	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLJ07	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLJ12	51	1	1.92%	0	0.00%	0	0.00%	2	3.92%
SZLJ13	18	34	65.38%	0	0.00%	0	0.00%	0	0.00%
SZBY08	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZMJ04	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
SZBK02	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBK03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZMJ05	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%

Table 12. Specific weeks found to have problems (Southern facilities 1-37)

Facility	Missed Week(s)	Zero Week(s)	Non-Matching Week(s)	Duplicate Week(s)	Duplicate Case Count Weeks
SZLJ08	1-42	-	-	-	-
SZLJ01	-	-	35, 36	-	-
SZBY07	-	-	14	-	-
SZLJ03	6-52	-	-	-	-
SZBY02	-	-	52	-	(33,34)
SZBY01	-	-	-	-	-
SZGE03	-	-	13, 30	-	-
SZGE02	-	-	23, 44, 52	-	-
SZLJ09	6-52	-	-	-	-
SZGE08	37-44	-	-	-	-
SZBY06	-	-	-	-	-
SZLJ02	-	-	18	-	-
SZGE04	17	-	26	-	(25,26)
SZGE11	-	-	-	-	(38,39),(23,24),(30,31)
SZGE06	-	-	-	-	-
SZGE07	10	-	20, 28, 52	-	-
SZLJ10	-	-	47	-	-
SZGE10	16	-	-	-	(45,46),(30,31)
SZGE05	-	-	-	-	(23,24)
SZGE01	-	-	-	-	-
SZLJ05	-	-	37	-	-
SZBK01	-	-	47	-	-
SZMJ01	-	-	30	-	-
SZLJ06	-	-	48	-	(6,39)
SZLJ11	-	-	49	-	(21,34)
SZLJ04	6-52	-	-	-	-
SZBY04	-	-	-	-	(1,2)
SZGE12	-	-	10	-	-
SZBY05	-	-	28, 47, 50	-	-
SZLJ07	-	-	-	-	-
SZLJ12	37	-	39, 50	-	(35,41),(47,48)
SZLJ13	6-39	-	49	-	-
SZBY08	-	-	-	-	-
SZMJ04	-	-	50	-	(35,37),(41,47)
SZBK02	-	-	-	-	-
SZBK03	-	-	-	-	-
SZMJ05	-	-	26, 30, 36	-	(11,30)

Table 13. Digit preference breakdown, Southern zone, χ^2/df values

facility	total cases 5 and over	total cases less than 5	total cases female	total cases male	other cases 5 and over	other cases less than 5	other cases female	other cases cases male	overall χ^2/df	overall p-value
SZLJ08	0.16	0.17	0.64	0.25	0.43	1.90	0.43	0.40	0.43	1.00
SZLJ01	1.79	0.89	0.72	0.68	1.79	3.45*	1.94	3.54*	1.82	<.01**
SZBY07	0.63	0.50	2.00	1.27	0.80	0.73	1.19	1.36	1.06	0.35
SZLJ03	0.20	0.00	0.00	0.20	0.00	0.20	0.00	0.20	0.09	1.00
SZBY02	0.63	1.27	0.97	0.55	0.59	1.40	0.97	1.06	0.93	0.64
SZBY01	2.00	0.46	2.43	1.32	0.97	1.19	2.26	0.85	1.43	0.01**
SZGE03	0.85	1.74	0.89	0.97	0.33	1.87	0.46	0.63	0.97	0.55
SZGE02	0.80	0.29	0.77	2.04	0.63	0.63	1.19	3.11*	1.19	0.13
SZLJ09	0.20	0.20	0.20	0.20	0.20	0.00	0.20	0.00	0.14	1.00
SZGE08	1.12	0.77	0.82	0.52	1.53	0.52	0.67	0.41	0.79	0.90
SZBY06	0.89	0.72	1.74	1.55	0.59	0.85	1.53	0.59	1.05	0.36
SZLJ02	0.68	0.72	1.44	0.89	1.62	2.67*	1.49	2.09	1.43	0.01**
SZGE04	2.15	0.80	2.20	0.50	1.89	1.28	1.50	1.54	1.48	<.01**
SZGE11	1.06	1.62	0.76	0.42	2.21	0.60	0.25	1.19	1.03	0.41
SZGE06	1.02	1.02	1.15	2.00	1.02	0.72	1.23	1.87	1.25	0.07
SZGE07	1.33	1.20	0.89	1.24	1.24	1.72	1.76	2.60	1.48	0.01**
SZLJ10	0.59	1.49	0.50	0.76	1.66	1.44	1.66	2.73*	1.35	0.02**
SZGE10	0.89	0.80	1.15	0.76	0.98	2.94*	1.33	1.76	1.33	0.03**
SZGE05	1.10	0.93	1.40	0.51	0.85	0.89	1.40	2.21	1.17	0.15
SZGE01	1.10	0.68	1.32	2.17	1.27	2.21	1.91	2.9*	1.66	0.00**
SZLJ05	0.89	1.06	0.68	0.97	1.66	2.09	1.59	2.34	1.42	0.01**
SZBK01	1.02	1.27	0.42	0.72	0.72	0.93	0.21	1.70	0.87	0.77
SZMJ01	0.85	0.63	0.46	0.77	0.46	1.53	0.63	1.12	0.80	0.89
SZLJ06	.	1.15	0.97	0.89	.	1.91	0.85	1.74	1.25	0.10
SZLJ11	0.42	0.97	0.55	1.44	1.53	0.80	0.72	1.15	0.95	0.60
SZLJ04	0.20	0.20	0.20	0.20	0.20	0.20	0.00	0.20	0.17	1.00
SZBY04	1.06	1.74	0.97	2.00	1.15	0.94	0.76	1.02	1.21	0.11
SZGE12	1.10	1.32	0.68	1.02	0.50	1.27	2.34	0.46	1.09	0.29
SZBY05	1.06	0.33	0.63	0.55	0.80	1.27	0.50	0.76	0.74	0.95
SZLJ07	1.23	0.63	0.63	0.59	0.68	0.85	0.50	0.33	0.68	0.98
SZLJ12	1.50	1.41	1.94	0.89	1.02	1.59	0.62	1.28	1.29	0.05
SZLJ13	0.88	1.80	0.63	0.50	0.86	0.47	0.73	0.35	0.73	0.94
SZBY08	0.76	1.15	0.76	1.02	1.44	1.70	0.68	0.50	1.00	0.48
SZMJ04	1.19	1.23	0.72	0.76	1.44	1.29	0.89	1.74	1.16	0.17
SZBK02	1.10	1.83	0.73	1.36	0.97	0.93	0.97	0.97	1.11	0.24
SZBK03	0.93	0.64	0.50	0.33	0.68	1.36	1.96	1.62	1.01	0.46
SZMJ05	1.74	1.57	1.32	0.72	3.37*	2.26	1.81	4.56*	2.17	<.01**

* value is significant at .05/8 level (Method 1)

**overall value is significant at .05 level (Method 2)

Table 14. Number of reports, number of missing reports, number of duplicate weeks, number of weeks with all zeros, and number of duplicate case count weeks in Central facilities (facilities 1-30)

Facility	Number of Reports	Total Missing Weeks	% Missing Weeks	Total Number of Weeks with all Zeros	% Weeks with all Zeros	Total Duplicate Weeks	% Duplicate Weeks	Duplicate Case Count Weeks	% Duplicate Case Count Weeks
SZBN11	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN01	53	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGA01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGA02	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN17	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS17	26	26	50.00%	0	0.00%	0	0.00%	1	3.85%
SZLS09	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLS14	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZMS01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN18	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS12	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLS13	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLS10	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN02	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZHA01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS11	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN23	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGA03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGA04	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGA05	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZGA06	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
SZLS01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZGA07	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS03	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
SZLS18	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN05	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN09	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SZBN10	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SZBN04	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%

Table 15. Number of reports, number of missing reports, number of duplicate weeks, number of weeks with all zeros, and number of duplicate case count weeks in Central facilities (facilities 31-61)

Facility	Number of Reports	Total Missing Weeks	% Missing Weeks	Total Number of Weeks with all Zeros	% Weeks with all Zeros	Total Duplicate Weeks	% Duplicate Weeks	Duplicate Case Count Weeks	% Duplicate Case Count Weeks
SZBN12	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS15	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZHA02	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLS19	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZMS05	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS20	50	2	3.85%	0	0.00%	0	0.00%	1	2.00%
SZBN16	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS16	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN24	1	51	98.08%	0	0.00%	0	0.00%	0	0.00%
SZLS06	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
SZBN28	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN20	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN21	50	2	3.85%	0	0.00%	0	0.00%	1	2.00%
SZLS21	44	8	15.38%	0	0.00%	0	0.00%	0	0.00%
SZBN29	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZMS06	14	38	73.08%	0	0.00%	0	0.00%	1	7.14%
SZLS22	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLS26	52	0	0.00%	0	0.00%	0	0.00%	4	7.69%
SZLS27	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN22	50	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN14	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS28	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZLS23	1	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS24	52	2	3.85%	0	0.00%	0	0.00%	1	1.92%
SZLS08	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN26	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZLS30	50	0	0.00%	0	0.00%	0	0.00%	2	4.00%
SZLS29	44	0	0.00%	0	0.00%	0	0.00%	2	4.55%
SZMS08	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SZBN30	14	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SZBN13	14	0	0.00%	0	0.00%	0	0.00%	0	0.00%

Table 16. Specific weeks found to have problems (Central facilities 1-30)

Facility	Missed Week(s)	Zero Week(s)	Non-Matching Week(s)	Duplicate Week(s)	Duplicate Case Count Weeks
SZBN11	-	-	-	-	-
SZBN01	-	-	*	-	-
SZGA01	-	-	30	-	-
SZGA02	-	-	7	-	-
SZBN17	-	-	4	-	-
SZLS17	7,28-52	-	-	-	(21, 26)
SZLS09	-	-	17, 48	-	(30, 31)
SZLS14	-	-	42	-	(30,31)
SZMS01	-	-	1, 4	-	-
SZBN18	-	-	15, 46	-	-
SZLS12	-	-	37	-	(30,31)
SZLS13	-	-	-	-	(30,31)
SZLS10	-	-	50	-	(30,31)
SZBN02	-	-	-	-	-
SZHA01	-	-	45	-	-
SZLS11	-	-	-	-	(30,31)
SZBN23	-	-	-	-	(13,14)
SZBN03	-	-	-	-	-
SZGA03	-	-	-	-	-
SZGA04	-	-	36	-	-
SZGA05	-	-	-	-	(12,13)
SZGA06	-	-	-	-	(6,7), (18,19)
SZLS01	-	-	-	-	-
SZGA07	-	-	43	-	-
SZLS03	-	-	51	-	(36,38), (17,18)
SZLS18	-	-	2	-	(30,31)
SZBN05	-	-	-	-	-
SZBN09	27	-	29	-	-
SZBN10	27	-	-	-	-
SZBN04	-	-	-	-	(2,24)

*week label missing

Table 17. Specific weeks found to have problems (Central facilities 31-61)

Facility	Missed Week(s)	Zero Week(s)	Non-Matching Week(s)	Duplicate Week(s)	Duplicate Case Count Weeks
SZBN12	-	-	-	-	-
SZLS15	-	-	-	-	(30,31)
SZHA02	-	-	9	-	(6,7)
SZLS19	-	-	-	-	(30,31)
SZMS05	-	-	-	-	-
SZLS20	51, 52	-	-	-	(30,31)
SZBN16	-	-	-	-	-
SZLS16	-	-	-	-	(30,31)
SZBN24	1, 3-52	-	-	-	-
SZLS06	-	-	36	-	(50,37), (22,26)
SZBN28	-	-	-	-	-
SZBN20	-	-	25	-	-
SZBN21	31, 32	-	-	-	(39,45)
SZLS21	28,29,30,45,46,49,51,52	-	1	-	-
SZBN29	-	-	-	-	-
SZMS06	15-52	-	-	-	(1,9)
SZLS22	-	-	27	-	(30,31)
SZLS26	-	-	-	-	(36,4), (40,46), (35,45), (6,7)
SZLS27	-	-	-	-	(38,39)
SZBN22	-	-	*	-	-
SZBN14	-	-	35, 37	-	-
SZLS28	-	-	-	-	(8,11)
SZLS23	-	-	7, 34	-	-
SZLS24	33,46	-	-	-	(30,31)
SZLS08	-	-	-	-	(45,46)
SZBN26	-	-	-	-	-
SZLS30	-	-	-	-	(30,31),(21,46)
SZLS29	-	-	-	-	(30,31), (49,50)
SZMS08	-	-	-	-	(3,4)
SZBN30	-	-	-	-	-
SZBN13	-	-	-	-	-

*week label missing

Table 18. Digit preference breakdown, Central zone (facilities 1-30), χ^2/df values

facility	total cases 5 and over	total cases less than 5	total cases female	total cases male	other cases 5 and over	other cases less than 5	other cases female	other cases male	overall χ^2/df	overall p-value
SZBN11	48.08*	0.33	0.97	0.80	48.08*	0.38	1.27	0.80	2.45	<.01**
SZBN01	0.76	1.27	0.50	1.06	0.59	1.15	0.72	1.10	0.89	0.73
SZGA01	1.23	1.32	1.36	0.93	1.23	0.97	1.02	2.73*	1.35	0.03**
SZGA02	0.33	0.59	1.15	0.89	0.76	0.94	1.36	1.10	0.89	0.73
SZBN17	0.81	48.08*	0.68	0.59	1.02	48.08*	1.53	0.76	2.61	<.01**
SZLS17	0.47	1.38	0.95	0.96	0.99	1.38	1.56	1.47	1.13	0.22
SZLS09	0.93	1.40	1.23	0.50	1.79	1.06	0.68	0.68	1.03	0.40
SZLS14	0.63	1.23	0.68	1.15	1.74	1.79	0.85	2.34	1.30	0.04**
SZMS01	1.74	2.51	1.23	0.85	2.30	2.73*	0.68	0.94	1.65	<.01**
SZBN18	0.38	0.80	1.15	1.15	1.10	0.68	0.80	0.59	0.83	0.85
SZLS12	1.74	1.06	1.23	0.72	1.83	1.62	1.40	0.68	1.28	0.05
SZLS13	1.06	1.66	1.06	0.80	1.15	2.60*	1.76	1.19	1.40	0.01**
SZLS10	1.16	0.59	0.89	1.40	1.10	1.19	0.55	0.76	0.95	0.59
SZBN02	1.66	1.02	2.67*	1.15	1.70	1.15	1.53	1.02	1.47	0.01**
SZHA01	1.15	1.32	0.89	1.23	1.10	1.44	2.09	1.74	1.37	0.02**
SZLS11	1.20	0.68	2.73*	2.20	1.12	0.63	2.30	1.54	1.54	<.01**
SZBN23	4.82*	0.85	1.06	1.19	1.32	0.50	1.87	0.51	1.53	<.01**
SZBN03	0.76	4.09*	0.59	0.59	0.59	4.61*	0.50	0.68	1.55	<.01**
SZGA03	1.40	1.85	0.72	0.99	1.32	1.76	1.02	0.81	1.23	0.10
SZGA04	1.10	1.96	2.98*	0.50	2.26	3.88*	2.07	2.17	2.10	<.01**
SZGA05	1.32	1.40	1.07	0.89	1.79	2.13	0.76	0.72	1.26	0.07
SZGA06	1.59	1.44	1.79	0.68	1.16	1.32	0.89	1.44	1.29	0.05
SZLS01	0.89	46.17*	1.74	5.08*	0.80	46.17*	2.72*	3.71*	5.55	<.01**
SZGA07	0.93	0.63	1.49	1.27	0.50	0.93	1.32	0.46	0.94	0.62
SZLS03	0.72	1.19	1.10	1.70	1.32	0.80	1.10	1.15	1.13	0.20
SZLS18	1.27	0.93	0.76	2.38	1.27	1.19	0.72	1.70	1.28	0.06
SZBN05	1.27	0.97	0.89	1.66	1.53	1.79	1.19	1.38	1.33	0.03**
SZBN09	0.80	1.10	1.37	0.80	1.94	1.07	1.50	0.50	1.13	0.21
SZBN10	3.33*	0.37	1.02	1.20	2.37	0.63	1.46	0.89	1.41	0.01**
SZBN04	1.32	1.06	0.94	1.89	1.32	1.06	0.94	1.89	1.28	0.06

* value is significant at .05/8 level (Method 1)

**overall value is significant at .05 level (Method 2)

Table 19. Digit preference breakdown, Central zone (facilities 31-61), χ^2/df values

facility	total cases 5 and over	total cases less than 5	total cases female	total cases male	other cases 5 and over	other cases less than 5	other cases female	other cases male	overall χ^2/df	overall p-value
SZBN12	1.10	0.42	0.50	1.44	0.93	0.93	0.81	1.74	0.99	0.50
SZLS15	0.93	0.63	0.21	1.15	0.89	2.81*	3.58*	4.01*	1.74	<.01**
SZHA02	1.23	1.23	1.10	0.72	1.66	1.44	0.59	1.15	1.14	0.20
SZLS19	0.97	2.09	0.33	0.80	1.32	3.20*	1.15	1.44	1.41	0.01**
SZMS05	1.06	0.29	1.23	0.38	0.72	0.80	0.55	0.76	0.72	0.96
SZLS20	1.02	1.02	0.80	0.67	0.89	1.29	1.07	0.58	0.92	0.68
SZBN16	0.42	1.87	1.70	0.68	1.23	1.19	1.27	0.63	1.12	0.22
SZLS16	1.15	0.86	1.06	1.27	0.93	1.02	0.81	2.38	1.20	0.13
SZBN24
SZLS06	1.44	2.33	1.49	0.59	1.10	1.89	1.23	0.72	1.33	0.03**
SZBN28	1.36	0.63	1.01	1.79	2.13	0.59	1.23	0.59	1.17	0.16
SZBN20	0.68	2.68*	0.80	1.42	0.89	1.36	0.76	0.50	1.13	0.21
SZBN21	1.11	34.10*	1.42	2.04	0.62	39.92*	0.93	0.62	5.20	<.01**
SZLS21	1.02	0.52	0.92	1.47	1.22	1.02	0.72	1.58	1.06	0.35
SZBN29	0.89	1.32	1.02	1.83	0.59	1.15	1.19	2.00	1.25	0.08
SZMS06	0.45	0.67	0.50	0.97	0.45	0.50	0.63	0.97	0.62	0.98
SZLS22	0.85	0.72	2.13	1.79	0.89	1.62	1.27	1.36	1.33	0.03**
SZLS26	1.32	3.28*	1.10	1.96	1.83	2.67*	0.97	1.44	1.81	<.01**
SZLS27	1.40	1.12	1.83	0.38	2.21	1.59	0.60	0.85	1.25	0.08
SZBN22	7.43*	9.19*	8.90*	11.03*	7.43*	9.19*	9.08*	11.03*	8.98	<.01**
SZBN14	1.40	0.76	1.32	0.46	1.15	0.76	0.55	1.10	0.94	0.63
SZLS28	0.38	0.25	1.36	0.68	0.42	0.21	1.44	1.06	0.72	0.96
SZLS23	0.85	1.06	0.55	1.70	0.97	0.89	1.83	0.89	1.09	0.28
SZLS24	0.93	0.27	0.67	1.02	0.71	1.07	0.62	1.11	0.80	0.89
SZLS08	1.74	1.87	1.15	1.15	2.00	1.74	1.27	1.23	1.52	<.01**
SZBN26	0.46	0.97	0.55	0.68	0.46	0.97	0.55	0.68	0.66	0.99
SZLS30	0.97	0.89	0.80	1.10	0.42	0.55	2.24	1.98	1.10	0.27
SZLS29	0.50	1.15	1.70	0.72	2.30	0.85	1.02	0.68	1.11	0.24
SZMS08	1.23	0.46	0.38	0.72	1.36	0.42	0.55	0.50	0.70	0.97
SZBN30	1.74	3.78*	1.15	0.89	1.62	4.99*	1.32	0.50	1.95	<.01**
SZBN13	1.32	2.21	0.80	0.59	1.19	2.21	0.89	0.38	1.20	0.12

* value is significant at .05/8 level (Method 1)

**overall value is significant at .05 level (Method 2)

Table 20. Number of reports, number of missing reports, number of duplicate weeks, number of weeks with all zeros, and number of duplicate case count weeks in Puntland facilities (facilities 1-25)

Facility	Number of Reports	Total Missing Weeks	% Missing Weeks	Total Number of Weeks with all Zeros	% Weeks with all Zeros	Total Duplicate Weeks	% Duplicate Weeks	Duplicate Case Count Weeks	% Duplicate Case Count Weeks
PLMU04	51	1	1.92%	0	0.00%	0	0.00%	2	3.92%
PLMU09	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLKA01	53	1	1.92%	0	0.00%	2	3.77%	0	0.00%
PLSA01	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
PLSA02	52	1	1.92%	1	1.92%	1	1.92%	5	9.62%
PLSL01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLBA03	52	0	0.00%	3	5.77%	0	0.00%	1	1.92%
PLBA01	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
PLBA02	53	0	0.00%	0	0.00%	1	1.89%	0	0.00%
PLCA01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLNG01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLBA04	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLNG04	53	0	0.00%	1	1.89%	1	1.89%	3	5.66%
PLNG03	54	0	0.00%	0	0.00%	2	3.70%	0	0.00%
PLSA04	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
PLKA02	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
PLMU05	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
PLMU03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLMU06	47	5	9.62%	0	0.00%	0	0.00%	0	0.00%
PLNG08	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
PLNG09	53	0	0.00%	1	1.89%	1	1.89%	0	0.00%
PLNG05	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
PLSA03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLKA03	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
PLMU07	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%

Table 21. Number of reports, number of missing reports, number of duplicate weeks, number of weeks with all zeros, and number of duplicate case count weeks in Puntland facilities (facilities 26-46)

Facility	Number of Reports	Total Missing Weeks	% Missing Weeks	Total Number of Weeks with all Zeros	% Weeks with all Zeros	Total Duplicate Weeks	% Duplicate Weeks	Duplicate Case Count Weeks	% Duplicate Case Count Weeks
PLSL02	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLBA05	52	2	3.85%	0	0.00%	2	3.85%	0	0.00%
PLNG02	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLMU10	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
PLNG10	50	2	3.85%	0	0.00%	0	0.00%	0	0.00%
PLSL03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLSL04	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
PLSL05	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLMU02	5	47	90.38%	0	0.00%	0	0.00%	0	0.00%
PLNG11	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLKA04	52	0	0.00%	1	1.92%	0	0.00%	0	0.00%
PLKA05	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLNG06	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLKA06	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLKA07	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
PLNG12	52	0	0.00%	1	1.92%	0	0.00%	1	1.92%
PLSL06	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
PLBA06	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
PLKA08	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PLNG07	53	1	1.92%	1	1.89%	2	3.77%	0	0.00%
PLSA05	52	0	0.00%	3	5.77%	0	0.00%	1	1.92%

Table 22. Specific weeks found to have problems (Puntland facilities 1-25)

Facility	Missed Week(s)	Zero Week(s)	Non-Matching Week(s)	Duplicate Week(s)	Duplicate Case Count Weeks
PLMU04	32	-	-	-	(19,29),(51,52)
PLMU09	-	-	-	-	-
PLKA01	13	-	-	12,26	-
PLSA01	-	-	-	-	(33,35),(15,16)
PLSA02	3	43	-	2	(29,31),(41,12),(8,51),15,16),(13,20)
PLSL01	-	-	-	-	-
PLBA03	-	2, 3, 4	-	-	(13,20)
PLBA01	35	-	47	-	(51,52)
PLBA02	-	-	18, 51	35	-
PLCA01	-	-	-	-	-
PLNG01	-	-	-	-	-
PLBA04	-	-	-	-	-
PLNG04	-	41	-	32	(14,24),(7,8),(25,28)
PLNG03	-	-	-	26,27	-
PLSA04	-	-	21	-	(28,30),(15,16)
PLKA02	-	-	-	-	(26,28),(46,47)
PLMU05	-	-	-	-	(40,46)
PLMU03	-	-	-	-	-
PLMU06	32,35,37,38,44	-	-	-	-
PLNG08	37	-	17, 26	-	-
PLNG09	-	37	-	37	-
PLNG05	-	-	-	-	(1,28)
PLSA03	-	-	-	-	-

Table 23. Specific weeks found to have problems (Puntland facilities 26-46)

Facility	Missed Week(s)	Zero Week(s)	Non-Matching Week(s)	Duplicate Week(s)	Duplicate Case Count Weeks
PLKA03	26	-	-	-	(46,47)
PLMU07	-	-	-	-	-
PLSL02	-	-	-	-	-
PLBA05	48,49	-	9	8,9	-
PLNG02	-	-	3, 48	-	-
PLMU10	-	-	-	-	(29,34)
PLNG10	26,27	-	-	-	-
PLSL03	-	-	-	-	-
PLSL04	-	-	-	-	(19,40)
PLSL05	-	-	-	-	-
PLMU02	1-31, 33, 34, 36, 39-43, 45-52	-	-	-	-
PLNG11	-	-	48	-	-
PLKA04	-	50	-	-	-
PLKA05	-	-	24, 52	-	-
PLNG06	-	-	5	-	-
PLKA06	-	-	-	-	-
PLKA07	-	-	-	-	(46,47)
PLNG12	-	48	-	-	(2,3)
PLSL06	-	-	-	-	(18,19)
PLBA06	-	-	-	-	(17,34)
PLKA08	-	-	-	-	-
PLNG07	48	52	-	8,52	-
PLSA05	-	33, 39, 40	-	-	(15,16)

Table 24. Digit preference breakdown, Puntland zone (facilities 1-25), χ^2/df values

facility	total cases 5 and over	total cases less than 5	total cases female	total cases male	other cases 5 and over	other cases less than 5	other cases female	other cases male	overall χ^2/df	overall p-value
PLMU04	2.85*	10.46*	0.98	0.79	3.16*	10.46*	1.02	0.79	3.71	<.01**
PLMU09	1.32	1.06	1.36	0.89	1.32	1.06	1.36	0.89	1.16	0.17
PLKA01	0.76	1.60	0.30	0.46	1.01	1.60	0.34	0.63	0.84	0.84
PLSA01	1.10	1.32	1.32	1.49	1.40	1.19	1.23	0.97	1.25	0.07
PLSA02	1.66	2.56*	0.80	1.38	1.62	2.77*	0.80	1.81	1.68	<.01**
PLSL01	1.06	0.80	2.00	2.81*	1.06	0.80	2.00	2.81*	1.67	<.01**
PLBA03	0.55	1.27	0.85	1.57	2.90*	1.83	0.29	1.23	1.31	0.04**
PLBA01	1.41	2.94*	1.02	0.85	1.59	2.46	1.59	1.94	1.72	<.01**
PLBA02	0.92	0.51	0.71	1.51	1.05	1.01	0.76	0.88	0.93	0.64
PLCA01	0.76	0.55	2.89*	1.94	0.76	0.55	2.89*	1.94	1.48	0.01**
PLNG01	0.46	0.73	0.68	1.15	0.50	0.93	0.85	1.62	0.87	0.78
PLBA04	1.02	1.06	1.49	0.72	0.80	0.89	1.10	0.63	0.96	0.57
PLNG04	13.10*	2.73*	0.88	2.18	13.10*	2.69*	1.09	1.76	4.52	<.01**
PLNG03	0.96	2.56*	1.08	1.33	1.46	1.82	1.00	1.53	1.47	0.01**
PLSA04	0.55	0.63	0.55	0.50	0.76	0.38	0.68	0.76	0.60	1.00
PLKA02	0.63	2.85*	0.80	0.50	0.76	2.85*	0.89	0.50	1.18	0.15
PLMU05	0.72	1.23	0.68	0.64	0.72	1.23	0.68	0.64	0.82	0.86
PLMU03	2.81*	3.20*	2.30	1.70	2.68*	3.11*	3.28*	2.09	2.65	<.01**
PLMU06	0.71	0.81	1.04	0.90	1.66	0.66	3.01*	1.00	1.20	0.12
PLNG08	0.67	1.37	0.19	0.93	0.59	1.20	0.41	1.11	0.81	0.88
PLNG09	0.71	0.71	0.84	0.17	0.71	1.01	1.26	0.51	0.74	0.95
PLNG05	6.40*	6.00*	0.76	1.02	6.40*	6.00*	0.76	1.02	3.48	<.01**
PLSA03	0.97	2.09	1.49	0.93	0.97	2.09	1.49	0.93	1.37	0.02**
PLKA03	1.68	0.50	0.63	0.93	1.68	0.50	0.63	0.93	0.93	0.64
PLMU07	0.59	1.27	1.44	0.80	0.59	1.27	1.44	0.80	1.03	0.41

* value is significant at .05/8 level (Method 1)

**overall value is significant at .05 level (Method 2)

Table 25. Digit preference breakdown, Puntland zone (facilities 26-46), χ^2/df values

facility	total cases 5 and over	total cases less than 5	total cases female	total cases male	other cases 5 and over	other cases less than 5	other cases female	other cases male	overall χ^2/df	overall p-value
PLSL02	1.66	1.49	1.10	2.09	1.66	1.49	1.10	2.09	1.58	<.01**
PLBA05	0.72	0.46	1.10	1.19	1.15	0.63	0.97	1.06	0.91	0.69
PLNG02	0.89	0.72	1.40	1.32	0.72	0.93	1.32	1.19	1.06	0.34
PLMU10	2.00	1.02	1.81	1.83	2.00	1.02	1.81	1.83	1.66	<.01**
PLNG10	1.87	0.64	1.56	1.31	1.42	1.60	1.47	1.38	1.42	0.01**
PLSL03	0.68	0.50	1.98	1.10	0.68	0.50	1.98	1.10	1.05	0.36
PLSL04	0.85	1.87	1.66	1.74	0.85	1.87	1.66	1.74	1.53	<.01**
PLSL05	0.93	1.76	3.75*	1.15	0.93	1.76	3.75*	1.15	1.85	<.01**
PLMU02	0.00	0.20	0.20	0.20	0.20	0.00	0.00	0.20	0.11	1.00
PLNG11	1.62	1.36	0.59	0.90	0.76	0.73	0.80	0.50	0.91	0.69
PLKA04	1.10	1.53	1.62	0.63	0.72	1.53	1.40	0.46	1.12	0.22
PLKA05	0.93	0.89	1.40	1.40	0.68	0.68	0.60	1.32	0.99	0.50
PLNG06	0.97	1.53	1.15	0.76	0.97	1.53	1.15	0.76	1.10	0.26
PLKA06	0.72	1.15	1.83	0.93	0.68	1.23	1.62	0.80	1.12	0.23
PLKA07	0.97	0.59	2.38	1.23	0.76	0.59	2.21	1.15	1.24	0.08
PLNG12	2.00	3.88*	0.55	0.72	2.26	3.54*	0.42	0.63	1.75	<.01**
PLSL06	0.38	1.91	1.27	0.55	0.38	1.91	1.27	0.55	1.03	0.41
PLBA06	1.10	1.19	1.15	1.40	0.80	1.19	1.15	1.40	1.17	0.15
PLKA08	2.77*	0.99	1.19	0.68	3.41*	0.99	0.80	0.42	1.42	0.01**
PLNG07	2.52	1.09	0.59	1.01	2.27	1.09	0.71	1.01	1.29	0.05
PLSA05	0.68	2.13	2.09	1.66	0.68	2.13	2.09	1.66	1.64	<.01**

* value is significant at .05/8 level (Method 1)

**overall value is significant at .05 level (Method 2)

Table 26. Number of reports, number of missing reports, number of duplicate weeks, number of weeks with all zeros, and number of duplicate case count weeks in Somaliland facilities (facilities 1-30)

Facility	Number of Reports	Total Missing Weeks	% Missing Weeks	Total Number of Weeks with all Zeros	% Weeks with all Zeros	Total Duplicate Weeks	% Duplicate Weeks	Duplicate Case Count Weeks	% Duplicate Case Count Weeks
SLMA06	51	1	1.92%	1	1.96%	0	0.00%	0	0.00%
SLTO01	54	1	1.92%	0	0.00%	3	5.56%	9	16.67%
SLMA07	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SLSN01	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
SLMA01	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
SLMA02	50	2	3.85%	0	0.00%	0	0.00%	1	2.00%
SLSN04	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
SLSH01	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SLSH04	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SLAW02	53	0	0.00%	0	0.00%	1	1.89%	0	0.00%
SLAW03	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SLTO02	51	2	3.85%	0	0.00%	1	1.96%	6	11.76%
SLTO03	50	2	3.85%	0	0.00%	0	0.00%	10	20.00%
SLTO04	51	1	1.92%	0	0.00%	0	0.00%	4	7.84%
SLAW04	52	0	0.00%	0	0.00%	0	0.00%	0	0.00%
SLMA08	52	1	1.92%	0	0.00%	1	1.92%	0	0.00%
SLSO03	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
SLMA09	50	3	5.77%	0	0.00%	1	2.00%	1	2.00%
SLSN02	52	0	0.00%	0	0.00%	0	0.00%	3	5.77%
SLAW01	52	1	1.92%	0	0.00%	1	1.92%	0	0.00%
SLTO05	50	2	3.85%	0	0.00%	0	0.00%	9	18.00%
SLTO06	51	2	3.85%	0	0.00%	1	1.96%	9	17.65%
SLSN03	51	1	1.92%	0	0.00%	0	0.00%	2	3.92%
SLSN05	53	1	1.92%	9	16.98%	2	3.77%	0	0.00%
SLSO04	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
SLTO07	50	2	3.85%	0	0.00%	0	0.00%	10	20.00%
SLMA03	51	1	1.92%	4	7.84%	0	0.00%	0	0.00%
SLMA04	51	2	3.85%	0	0.00%	1	1.96%	2	3.92%
SLAW07	51	2	3.85%	0	0.00%	1	1.96%	0	0.00%
SLAW05	51	2	3.85%	0	0.00%	1	1.96%	0	0.00%

Table 27. Number of reports, number of missing reports, number of duplicate weeks, number of weeks with all zeros, and number of duplicate case count weeks in Somaliland facilities (facilities 31-54)

Facility	Number of Reports	Total Missing Weeks	% Missing Weeks	Total Number of Weeks with all Zeros	% Weeks with all Zeros	Total Duplicate Weeks	% Duplicate Weeks	Duplicate Case Count Weeks	% Duplicate Case Count Weeks
SLMA10	52	1	1.92%	0	0.00%	1	1.92%	0	0.00%
SLMA11	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SLMA12	50	2	3.85%	0	0.00%	0	0.00%	0	0.00%
SLSO01	49	3	5.77%	0	0.00%	0	0.00%	0	0.00%
SLMA13	52	1	1.92%	0	0.00%	1	1.92%	0	0.00%
SLAW08	52	1	1.92%	0	0.00%	1	1.92%	0	0.00%
SLSH02	53	0	0.00%	0	0.00%	1	1.89%	0	0.00%
SLSO02	51	1	1.92%	0	0.00%	0	0.00%	2	3.92%
SLTO08	51	1	1.92%	0	0.00%	0	0.00%	8	15.69%
SLTO09	51	1	1.92%	0	0.00%	0	0.00%	12	23.53%
SLSN06	52	0	0.00%	0	0.00%	0	0.00%	2	3.85%
SLSO05	50	2	3.85%	0	0.00%	0	0.00%	1	2.00%
SLSO06	53	1	1.92%	0	0.00%	2	3.77%	1	1.89%
SLMA14	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SLMA15	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SLTO10	51	1	1.92%	0	0.00%	0	0.00%	7	13.73%
SLSH03	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SLMA16	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
SLMA17	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SLAW06	52	0	0.00%	0	0.00%	0	0.00%	1	1.92%
SLSH05	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
SLSO08	51	1	1.92%	0	0.00%	0	0.00%	1	1.96%
SLMA05	51	1	1.92%	0	0.00%	0	0.00%	0	0.00%
SLSO07	51	1	1.92%	1	1.96%	0	0.00%	4	7.84%

Table 28. Specific weeks found to have problems (Somaliland facilities 1-30)

Facility	Missed Week(s)	Zero Week(s)	Non-Matching Week(s)	Duplicate Week(s)	Duplicate Case Count Weeks
SLMA06	41	1	18	-	-
SLTO01	41	-	22, 27	4, 22, 25	(29,50),(5,38),(16,40),(14,30),(13,32),(10,11,12),(7,31,42)
SLMA07	41	-	6, 9, 21, 26	-	-
SLSN01	-	-	28	-	(42,43),(49,50)
SLMA01	41	-	21	-	(45,50)
SLMA02	41, 43	-	-	-	(14,26)
SLSN04	-	-	-	-	(42,43),(49,50)
SLSH01	-	-	-	-	-
SLSH04	-	-	6, 26, 51	-	-
SLAW02	-	-	22	43	-
SLAW03	-	-	17,34	-	-
SLTO02	25, 41	-	48	51	(29,50),(5,38),(13,32),(10,11,12),(16,40)
SLTO03	41, 51	-	-	-	(5,38),(22,39),(14,30),(13,32),(7,31,42),(16,40),(20,35),(10,11,12)
SLTO04	41	-	20, 36, 48	-	(5,6),(10,11,12),(13,32),
SLAW04	-	-	-	-	-
SLMA08	41	-	2, 18, 22, 36	6	-
SLSO03	41	-	4, 18, 42	-	(33,34)
SLMA09	19,24,41	-	16	13	(30,39)
SLSN02	-	-	21	-	(42,43),(24,25),(49,50)
SLAW01	37	-	12,39	27	-
SLTO05	17, 41	-	-	-	(5,38),(13,32),(20,35),(14,30),(10,11,12),(29,50),(51,52),(7,31)
SLTO06	11, 41	-	-	37	(29,50),(22,39),(2,5,3),(14,30),(16,40),(13,32),(10,12),(7,31)
SLSN03	39	-	-	-	(42,43),(49,50)
SLSN05	52	49, 48, 43, 42, 29, 40, 41, 38, 50	2, 4	17, 30	-
SLSO04	41	-	49	-	(33,34)
SLTO07	37,41	-	30	-	(5,38),(10,11,12),(51,52),(22,39),(47,48),(31,42,6,7),(13,32)
SLMA03	41	6, 7, 8, 9	24, 28, 49, 52, 30	-	-
SLMA04	41, 3	-	-	1	(17,18),(8,13)
SLAW07	11, 48	-	-	45	-
SLAW05	26, 48	-	30	45	-

Table 29. Specific weeks found to have problems (Somaliland facilities 31-54)

Facility	Missed Week(s)	Zero Week(s)	Non-Matching Week(s)	Duplicate Week(s)	Duplicate Case Count Weeks
SLMA10	41	-	-	24	-
SLMA11	41	-	7, 9, 28	-	-
SLMA12	41, 47	-	22	-	-
SLSO01	4, 22, 41	-	21	-	(33,34)
SLMA13	41	-	6	47	-
SLAW08	6	-	16	11	-
SLSH02	-	-	-	11	-
SLSO02	41	-	20	-	(24,35),(33,34)
SLTO08	41	-	1	-	(29,50),(10,11,12),(7,31,42),(14,30),(20,35),13,32)
SLTO09	41	-	-	-	(3,5,38),(51,52),(29,50),(6,7,31),(10,11,12),(8,9),(13,32),(14,30), (20,35)
SLSN06	-	-	35	-	(42,43),(49,50)
SLSO05	24, 41	-	2,7,14,18	-	(33,34)
SLSO06	41	-	26	11, 24	(33,34)
SLMA14	41	-	-	-	-
SLMA15	41	-	17	-	-
SLTO10	41	-	-	-	(11,12),(9,10),(13,32),(5,38),(14,30),(7,31,42)
SLSH03	11	-	-	-	-
SLMA16	41	-	17	-	(34,35)
SLMA17	41	-	-	-	-
SLAW06	-	-	-	-	(12,49)
SLSH05	17	-	-	-	(14,15)
SLSO08	41	-	1,2	-	(33,34)
SLMA05	41	-	7, 17, 19, 30	-	-
SLSO07	41	10	-	-	(22,24,11),(7,25),(33,34)

Table 30. Digit preference breakdown, Somaliland zone (facilities 1-26), χ^2/df values

facility	total cases 5 and over	total cases less than 5	total cases female	total cases male	other cases 5 and over	other cases less than 5	other cases female	other cases male	overall χ^2/df	overall p-value
SLMA06	0.80	2.72*	0.76	0.76	0.80	2.72*	0.76	0.59	1.24	0.08
SLTO01	1.41	2.85*	2.85*	1.04	1.49	2.48	3.34*	1.37	2.10	<.01**
SLMA07	1.24	0.63	0.67	0.54	1.37	0.72	0.67	0.67	0.81	0.87
SLSN01	1.19	0.93	1.40	1.23	1.02	1.23	0.80	0.76	1.07	0.32
SLMA01	0.32	0.80	0.67	1.15	0.32	1.02	0.67	1.11	0.76	0.94
SLMA02	1.60	0.67	1.47	0.89	1.56	0.62	1.29	0.89	1.12	0.22
SLSN04	0.42	0.55	0.63	0.68	1.40	0.55	0.93	0.89	0.76	0.93
SLSH01	1.49	0.46	0.72	0.59	1.49	0.46	0.72	0.59	0.81	0.87
SLSH04	0.76	1.15	0.50	0.89	0.89	0.93	0.76	0.63	0.81	0.87
SLAW02	1.51	0.71	0.84	1.09	1.93	0.63	0.84	1.01	1.07	0.32
SLAW03	1.91	0.38	1.27	1.27	2.60*	1.19	1.49	1.74	1.48	<.01**
SLTO02	1.46	1.68	4.01*	1.76	1.20	1.85	4.99*	1.94	2.30	<.01**
SLTO03	1.81	2.00	1.27	3.38*	2.04	1.64	1.45	3.51*	2.16	<.01**
SLTO04	2.94*	0.11	1.76	2.11	1.59	0.41	3.94*	2.06	1.86	<.01**
SLAW04	2.34	0.80	1.06	0.76	1.66	0.85	1.23	1.10	1.23	0.09
SLMA08	0.59	0.59	0.46	0.80	0.63	0.85	0.33	0.59	0.61	1.00
SLSO03	0.54	1.98	1.89	1.68	0.41	2.42	1.68	1.32	1.49	<.01**
SLMA09	0.84	1.29	2.53	1.56	1.02	1.60	3.02*	2.09	1.74	<.01**
SLSN02	1.40	0.72	0.50	0.46	1.23	0.72	0.68	0.76	0.81	0.88
SLAW01	0.76	1.10	0.89	0.50	0.76	1.10	0.93	0.38	0.80	0.89
SLTO05	1.07	1.20	1.29	1.42	1.07	1.20	1.29	1.42	1.24	0.08
SLTO06	1.63	2.46	0.53	1.07	1.63	3.51*	0.79	1.07	1.61	<.01**
SLSN03	1.11	1.37	0.76	0.59	1.11	1.33	0.63	0.59	0.93	0.64
SLSN05	1.34	4.91*	1.39	2.35	0.88	5.08*	1.72	2.27	2.49	<.01**
SLSO04	1.20	1.01	1.07	1.28	1.28	1.37	1.50	1.07	1.23	0.10
SLTO07	4.23*	7.30*	3.33*	3.73*	4.27*	7.03*	3.11*	3.79*	4.54	<.01**

* value is significant at .05/8 level (Method 1)

**overall value is significant at .05 level (Method 2)

Table 31. Digit preference breakdown, Somaliland zone (facilities 27-54), χ^2/df values

facility	total cases 5 and over	total cases less than 5	total cases female	total cases male	other cases 5 and over	other cases less than 5	other cases female	other cases male	overall χ^2/df	overall p-value
SLMA03	0.85	1.11	0.85	1.11	2.15	1.33	0.80	1.07	1.16	0.17
SLMA04	0.80	0.67	0.63	1.54	0.80	0.67	0.63	1.54	0.91	0.69
SLAW07	0.85	1.20	0.32	0.85	0.85	1.46	0.41	0.85	0.85	0.82
SLAW05	0.85	1.59	1.50	1.41	0.85	1.50	1.54	1.33	1.32	0.04**
SLMA10	1.19	0.46	0.76	0.81	1.32	0.42	0.46	0.94	0.79	0.90
SLMA11	1.33	2.29	0.72	0.98	0.50	0.72	1.72	0.50	1.09	0.28
SLMA12	1.20	0.31	1.11	0.71	1.33	0.49	0.62	0.84	0.83	0.85
SLSO01	1.20	0.43	2.70*	0.70	1.15	0.29	2.20	0.61	1.16	0.17
SLMA13	0.89	0.85	1.02	1.91	1.23	0.85	1.19	1.87	1.23	0.09
SLAW08	1.32	1.10	0.29	0.89	2.09	0.60	0.42	0.89	0.95	0.59
SLSH02	0.57	0.51	0.88	0.97	0.57	0.42	0.88	1.09	0.74	0.95
SLSO02	0.80	1.50	1.19	1.24	0.67	1.98	1.33	1.02	1.22	0.10
SLTO08	2.02	2.42	0.98	2.24	1.54	2.68*	1.28	2.42	1.95	<.01**
SLTO09	1.59	2.02	4.68*	2.69	1.59	2.02	4.24*	2.69	2.67	<.01**
SLSN06	0.63	1.02	1.70	1.53	0.63	0.97	1.53	1.53	1.19	0.13
SLSO05	1.56	1.60	1.82	1.96	1.11	1.64	2.31	2.89*	1.86	<.01**
SLSO06	1.09	0.97	1.30	0.59	0.97	0.80	1.64	0.51	0.98	0.52
SLMA14	1.02	1.07	0.67	0.54	1.28	1.41	0.80	0.28	0.89	0.74
SLMA15	0.67	1.20	0.46	1.10	0.50	1.46	1.02	1.37	0.97	0.55
SLTO10	1.28	3.31*	1.63	2.59*	1.28	3.18*	1.81	2.59*	2.18	<.01**
SLSH03	1.94	0.80	1.98	1.63	1.94	1.15	1.46	1.41	1.54	<.01**
SLMA16	1.76	0.93	0.67	1.41	1.89	0.85	1.11	0.59	1.15	0.18
SLMA17	1.41	0.85	0.76	0.89	1.06	1.33	0.54	1.11	0.99	0.49
SLAW06	1.87	0.93	1.40	0.38	1.74	0.89	1.40	0.50	1.14	0.20
SLSH05	0.72	0.59	1.72	1.24	0.63	0.46	1.72	1.02	1.01	0.45
SLSO08	1.11	0.63	1.15	0.89	0.93	1.11	1.37	0.85	1.01	0.46
SLMA05	1.20	1.20	2.02	1.50	0.93	1.28	1.54	2.15	1.48	0.01**
SLSO07	1.54	3.03*	1.85	1.41	1.37	3.46*	1.85	1.07	1.95	<.01**

* value is significant at .05/8 level (Method 1)

**overall value is significant at .05 level (Method 2)

References

- 1) "Civil-Military Guidelines & Reference for Complex Emergencies." (2008): 8. *Refworld*. UN Office for the Coordination of Humanitarian Affairs, 2008. Web. 30 Mar. 2015.
- 2) Gargano, Lisa M., Jacqueline E. Tate, Umesh D. Parashar, Saad B. Omer, and Susan T. Cookson. "Comparison of Impact and Cost-effectiveness of Rotavirus Supplementary and Routine Immunization in a Complex Humanitarian Emergency, Somali Case Study." *Conflict and Health*. BioMed Central, 9 Feb. 2005. Web. 25 Mar. 2015.
- 3) Kouadio, Isidore K., Taro Kamigaki, and Hitoshi Oshitani. "Measles Outbreaks in Displaced Populations: A Review of Transmission, Morbidity and Mortality Associated Factors." *BMC International Health and Human Rights* 10.1 (2010): 5. Web.
- 4) Rajabali, Alefiyah, Omer Moin, Amna S. Ansari, Mohammad R. Khanani, and Syed H. Ali. "Communicable Disease among Displaced Afghans: Refuge without Shelter." *Nature Reviews Microbiology* 7.8 (2009): 609-14. Web.
- 5) "Military." *Somalia Civil War*. Globalsecurity.org, 5 Oct. 2013. Web. 24 Mar. 2015.
- 6) Kron, Josh, and Jeffrey Gettleman. "Kenya Says Western Nations Join Fight in Somalia, as U.S. Denies Role." *The New York Times*. The New York Times, 23 Oct. 2011. Web. 25 Mar. 2015.
- 7) "E. African Nations Back Kenyan Offensive in Somalia." *Voice of America*. N.p., 21 Oct. 2011. Web. 25 Mar. 2015.
- 8) *Eastern Africa Drought Humanitarian Report*. Rep. no. 3. OCHA, 10 June 2011. Web. 25 Mar. 2015.
- 9) "Somalia Famine 'killed 260,000'" *BBC News*. N.p., 2 May 2013. Web. 23 Mar. 2015.
- 10) *Analysis of the 2011 Communicable Diseases Surveillance Data*. Rep. WHO, n.d. Web. 25 Mar. 2015.
- 11) "Public Health Surveillance: A Tool for Targeting and Monitoring Interventions." *The Burden of Diseases in Resource Poor Countries: Meeting the Challenges of Combating HIV/AIDS, Tuberculosis, and Malaria*. Dar Es Salaam: National Institute for Medical Research, 2005. Print.
- 12) "Anthropometric Survey." *Anthropometric Survey*. SMART Methodology, 2015. Web. 19 Mar. 2015. <<http://www.nutrisurvey.de/ena/ena.html>>.
- 13) Centers for Disease Control and Prevention. Notes from the field: Malnutrition and mortality --- southern Somalia, July 2011. *MMWR* 2011; **60**(30): 1026-27

- 14) Casella, George, and Roger L. Berger. "Properties of a Random Sample." *Statistical Inference*. Australia: Thomson Learning, 2002. 219+. Print.
- 15) Taboga, M. (2010) "Lectures on probability and statistics", <http://www.statlect.com>.
- 16) Cox, D. R., and Peter A. W. Lewis. "Stationary Point Processes." *The Statistical Analysis of Series of Events*. London: Methuen, 1966. 72-73. Print.
- 17) Frank A. Haight (1967). *Handbook of the Poisson Distribution*. New York: John Wiley & Sons
- 18) Land, K. C., P. L. Mccall, and D. S. Nagin. "A Comparison of Poisson, Negative Binomial, and Semiparametric Mixed Poisson Regression Models: With Empirical Applications to Criminal Careers Data." *Sociological Methods & Research* 24.4 (1996): 387-442. Web.
- 19) Box, George E. P., and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Rev. ed. San Francisco: Holden-Day, 1976. Print.