ZIMBABWE ELECTION SUPPORT NETWORK

Promoting Democratic Elections in Zimbabwe

A comparative Analysis of the Voters Roll of Mutoko North Constituency

© ZESN November 2018

Executive Summary

The Mutoko North Constituency by-election will be held on the 24th of November 2018. As part of its oversight role on electoral processes and elections, the Zimbabwe Election Support Network (ZESN), conducted a computer audit of the Mutoko North Constituency voters roll, to provide an independent assessment of the quality of the register. The audit encompassed assessing accuracy, currency/ validity and comprehensiveness / completeness of the voters roll. This report thus presents the outcome of the computer audit.

A number of tests were conducted in accordance with an established methodology. The voters roll was compared to the 2018 Final Voters Roll (FVR), the 2013 Preliminary Voters Roll (PVR) and census data.

Overall the audit shows that there is a progressive improvement in the quality of the current roll when compared to the rolls used in previous elections especially the 2013 PVR.

The database used to store the data and the application used to capture the data appear to be having validation checks as there are very few mistakes in the capture of records. However, the fact that there is a duplicated ID number in the voters roll points to there being no unique constraint or at least a mechanism to alert data entry staff of the existence of duplicate records. There are problems regarding the accuracy of the dates of birth of individuals. This inaccuracy makes identification of multiple registered voters using biographical information more difficult.

There are no significant changes in the data presented for use in the 24th November 2018 by-election when compared to the data used in the 30th July 2018 election. This can be attributed to the short period of time between the two events. However, while the few changes recorded allude to the fact that there is an active process to add, update, remove and transfer voters– it's not clear when biometric deduplication is executed. The case of the duplication of TONNY ZISENGWE (48-049846 -S48) is evidence that the process is not run as often. The process of de-duplication should be run before each electoral event.

Finally, while the roll is largely representative of the population of Mutoko North since the differences between the actual registered voters and the projected registered voters is not significant – the Zimbabwe Electoral Commission (ZEC) should strive to ensure that they reduce the gap between the actual population and the registered voters.

Recommendations

Based on the findings from the computer audit, the ZESN makes the following recommendations:

- The ZEC should ensure consistency in capturing addresses of registrants. The ZEC should refer to the Alpha list to maintain consistency.
- There is need to accurately capture date of birth of registered voters.
- Deduplication should be done often, in order to identify duplicates. The ZEC should provide stakeholders with a detailed report of the deduplication process.
- The ZEC should display the voters roll outside inspection centres for identification of non-residents and the deceased.

Methodology

A computer audit is designed to detect potential problems with the registry. Because of the efficiency of this type of testing it is recommended that election management bodies ensure that these tests are conducted on an annual basis in order to help measure whether the quality of the register is improving, degrading or staying the same.

When done often voters roll audits can be valuable in increasing confidence in the register, or in case of errors and exceptions can alert election managers to the type and scope of such errors and inconsistencies and can indicate the time and level of effort required to make corrections. While many data integrity issues can be corrected with simple update queries some may require a field study to validate the suspicions raised by the audit.

The objectives of this particular audit are:

- To independently verify the quality of the Mutoko North Constituency voters roll
- To increase trust and confidence in the integrity of the Mutoko North voters roll

This is done by ascertaining or quantifying 3 aspects of the data in Mutoko North's registered voters - namely:

- The Accuracy of the data captured
- Currency of data
- Comprehensiveness of the data

All ages of voters in Mutoko North found in the rolls will be calculated as they will be on 24th November 2018.

To Determine Accuracy

Accuracy refers to the rate of error in the data. Data can be shown to be inaccurate in any of a number of ways while using data tests, namely:

1. The data is not possible (meaning that the value entered is not possible - an example of this would be a date of birth recorded as 29 Feb. 1990. Since 1990 was not a leap year, it's possible to establish that 29th of February 1990 is an invalid date - this can be identified by computer tests);
2. Missing data
3. Invalid data: records that should not be there in the registry. For example, persons who have registered multiple times, persons who are not citizens i.e. don't have valid ID numbers. The determination of whether those below 18 years have been registered also falls within this category of testing.

The extent to which these errors affect the credibility of an election varies from serious challenges that may disenfranchise voters or minor errors that may end up being an inconvenience to the voter and the polling station officials. For example, depending on polling day procedures, having a wrong ID number captured against a record may disenfranchise an

eligible voter. While having a special character in the name may break the sort order and inconvenience the voter when their names are being looked for in the register. The tests to be conducted will focus on;

1) Compliance to the legal framework
    a. Determination of underage persons i.e. under 18 years of age
    b. ID numbers will be checked for uniqueness and possibility of having duplicates.
    c. Testing the accuracy of the BVR deduplication process by doing a deduplication using biographical data. The approach that will be used to determine whether two or more records are to be regarded as belonging to the same individual shall involve grouping of voter records using the similarity of the fields within the voters' database.
2) Looking at the data fields captured and validating their accuracy by checking whether the data fields carry records that are probable. For example, names will be checked for the existence of numbers and special characters like "**: `~! @#$%^&***". The analysis will also check if voters have acceptable data in fields such as gender.
3) Some data fields such as Names shall be tested for validity based on uniqueness and length to determine if there are rows that are wrongly captured.
4) Dates of birth will be checked for existence (and validity if this is not enforced by the data source itself),
5) Days of birth will be checked for distribution across the days of the year. While it's quite probable that a substantial number of people will have unknown dates of birth and thus may state that their dates of birth were on the 1st of January on the year of their birth. Tests to determine the distribution of the dates of birth and gauge how well the voters are distributed will be executed. Sometimes the existence of irregular registration of voters can be detected by the distribution of these dates.

To Determine the Currency of Data - Increase/decrease of voting population over time

Currency refers to how up to date the data under scrutiny is. There are three ways in which this can be addressed

1) First missing youth are a sign of an out of date voters roll (i.e. people who have become eligible to vote since the last election). An age analysis will be used to show this.
2) Large numbers of very old (unrealistically old) people again suggest that the voters roll is not current (or at least one aspect of the update process is not working) because they have not been removed
3) Finally, a comparison of the current voters roll to be used for the by-election against the voters rolls from the last election (2018) and the 2013 PVR will be used to assess currency.

The tests that shall be conducted to ascertain currency are:

1) A voters roll demographic analysis to show how people are distributed. Barring irregularities, an age analysis of voters should result in a somewhat bell-shaped curve,

i.e. lower number of 18 and19-year old registrants, maximum number of voters aged in their 30'sand significant tapering off from age 50 onward. Similarly, gender ratios should remain more or less consistent across all wards. The Demographic analysis will detect any anomalies in the age/gender distribution of voters.

2) Attempting to establish whether there have been any changes over the life time of the register, and if so what these changes are. This will entail establishing the number of persons who were:
    a. added
    b. updated
    c. removed

    from the 2013 PVR and 2018 voters register.
3) An analysis of the number centenarians and persons between 18 and 19 gives a glimpse into the effectiveness of the process of inclusion of voters as well as removal of deceased persons.

To Determine Comprehensiveness

Comprehensiveness/completeness of a voters' register refers to the degree to which the voters roll contains all of the persons it should and the degree to which it includes ineligible voters. In order to establish completeness/comprehensiveness of the register, a comparison will be made between the population distribution by age group and sex data available in the census data and compared to the analysis from the register at the same level. The analysis will seek to find out:

1) The numbers of the underrepresented persons in Mutoko North Constituency.
2) Overrepresented persons in the Constituency
3) Compute the probable eligible population.
4) Compute and compare the gender ratio between the voters' register and the eligible population.

This process will not only give us a fair understanding of how comprehensive the register is but will also add value in the determination of currency especially the aspect of the removal of deceased/migrated persons and also how well it captures people who have recently come of voting age.

The margin of error in this case will be dependent on the methodology used by the ZESN to compute the population vis-à-vis the projection methodology that will be used to calculate the ages of eligible voters in Mutoko North.

Accuracy

Missing Data

The dataset provided had 11 pieces of information and it's expected that each registrant in the database should have:1) serial number 2) surname, 3) forenames, 4) national ID number, 5) gender, 6) date of birth, 7) address, 8) polling station code, 9) polling station name, 10) ward and 11) local authority. All the registrants in the dataset provided have the required fields.

The voting address is a free text field that seems to be in the [VILLAGE/HOUSE, HEADMAN, CHIEF, and DISTRICT] or something close to this. As seen in Table 1 below.

Table 1

Ser	Surname	Forenames	ID Number	Gender	Date of Birth	Voter Address
24	CHIKEYA	COLLEN	48-129574 - D48	Male	3/15/1999	KAMUKOSI VILLAGE, HEADMAN KAMUKOSI, CHIEF CHAREHWA
71	DIKE	ENOCK	48-027347 - M48	Male	10/1/1942	KUDYARAWANZA, HEADMAN CHIREKA, CHIEF CHAREHWA
91	HODZI	CARREN	48-175418 -J48	Male	7/26/1997	CHIREKA VILLAGE, HEADMAN DANDA, CHIEF CHAREHWA

However, some of the voting addresses seem to be missing parts of the data and these sections have been replaced with a "-". These records are as seen in Table 2 below.

Table 2

Ser	Surname	Forenames	ID Number	Gender	DOD	Voter Address
83	GEJO	FORGET	48-095576 - Z48	Male	11/28/1982	CHIREKA PRE-SCH, -, CHIEF CHAREWA, MUTOKO
2525	MATAMBANADZO	RUTENDO	48-153223 - J48	Female	3/5/1997	-, -, CHIEF CHAREWA, MUTOKO
697	MASUKU	DECENT	23-099963 - E23	Male	8/10/1992	-, -, CHIEF CHAREWA, MUTOKO
21687	MADHONGI	PEDZISAI	48-148752 - Z48	Male	2/16/1983	GONYE VILLAGE, -, CHIEF MUTOKO, MUTOKO
6625	MUCHANZA	SHINGIRAI	48-170228 - S49	Male	9/25/1998	SAMATANDA 2 VILLAGE, -, CHIEF CHAREWA, MUTOKO
21362	SHUMBA	SELUSIWE	03-114654 - S03	Female	5/11/1981	ZRP MUTOKO, -, CHIEF MUTOKO, MUTOKO

The ZEC should work on ensuring that addresses are correct and in the expected format.

Unclean Data

The tests to look for special characters i.e. ":;.`~!@#$%^&*"in the names of registrants did not return any records. All the records of registered voters also have ID numbers that are 14 characters long.

Dates of birth

All registrants have valid dates of births.

Under-age registrants

There are no underage voters.

Suspected Invalid Data

The day of birth analysis shows that there is an unnatural distribution of births across the calendar year. The two tables below exhibit this phenomenon. The top 15 days that have the most persons are listed in Table 3.Table 4gives the list of 15 days that have the least numbers of voters.

Table 3

Day of Birth	Voters
1-Jan	256
6-Jun	190
2-Feb	162
3-Mar	150
15-Jun	129
10-Jun	125
10-Mar	124
1-Jun	124
5-Jun	118
10-Jul	114
4-Apr	113
12-Jun	113
12-Apr	110
18-Apr	108
15-Mar	106

Table 4

Day of Birth	voters
Feb-29	11
31-Mar	20
30-Jul	23
31-Oct	23
31-May	28
30-Mar	28
30-May	29
29-Jan	29
30-Nov	30
23-Jan	30
8-Jul	31
29-Aug	33
31-Dec	33
31-Jan	33
28-Dec	33

This may be a result of voters not knowing their exact dates of birth i.e. 1st Jan (the beginning of the year) or 1st of June (mid-year). The large numbers of the rest of the days especially the ones that have the day of month being the same as the month itself i.e. 2nd of Feb, 3rd of March, 4th of April etc. juxtaposed next to those with fewest number of records (which are overwhelmingly more difficult days to write) – implies that there are cases where wrong dates of birth are being provided to the registration officials. While there are exceptions in this case

such as the 29th of February – that comes every 4 years – most of the other dates show lack of proportionality.

The chart in Figure 1 below plots these dates and as seen the outliers (on the right of the chart) point to a problem with the correctness of these data fields.

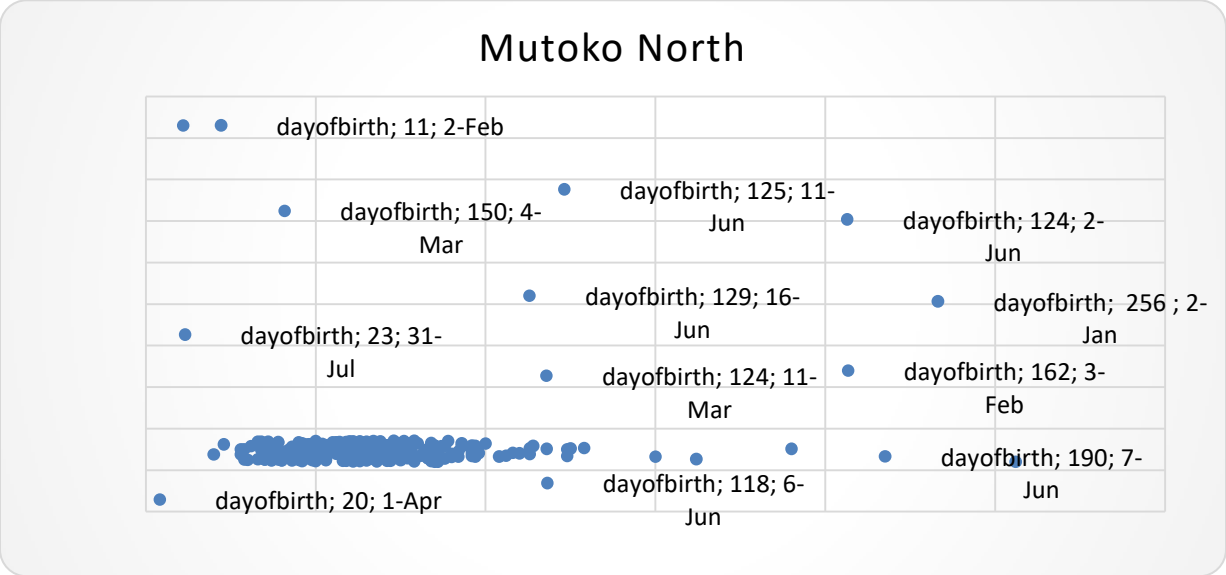

Figure 1

Suspected Multiple Registrations

The Roll provided has one suspected case of multiple registrations. This case is seen below:

Surname	Forenames	ID_Number	Gender
ZISENGWE	TONNY	48-049846 -S48	Male

Ser	DOB	Voter_Address	Polling_Station_Code
7963	7/11/1952	MACHONA VILLAGE, HEADMAN KAWERE, CHIEF MUTOKO, MUTOKO	4800MTK0402
8293	7/11/1957	MACHONA VILLAGE, KAWERE, MUTOKO	4800MTK0403

Currency

Existence of Young People in the Roll

The age-gender breakdown of the registered voters in Mutoko North is as shown in Table 5 below.

Table 5

Age Groups	Male	Female	Grand Total
18 - 19	352	294	646
20 - 24	1,400	1,398	2,798
25 - 29	1,178	1,331	2,509
30 - 34	1,267	1,571	2,838
35 - 39	1,291	1,491	2,782
40 - 44	1,018	1,303	2,321
45 - 49	939	1,055	1,994
50 - 54	527	726	1,253
55 - 59	413	814	1,227
60 - 65	557	1,013	1,570
65+	1,282	2,075	3,357
Grand Total	**10,224**	**13,071**	**23,295**

The roll has a total of 646 registrants that will be under 20 years old on 24th November 2018. This means that it has been updated recently to include them. However, none of these under 20s were added during the period of time between the publishing of the final roll and the extraction of this data.

Registrants over 100 years old

The age-gender chart (Figure 2) and table (Table 6) shown below shows that there may be a problem with removals of persons who are older than 100.

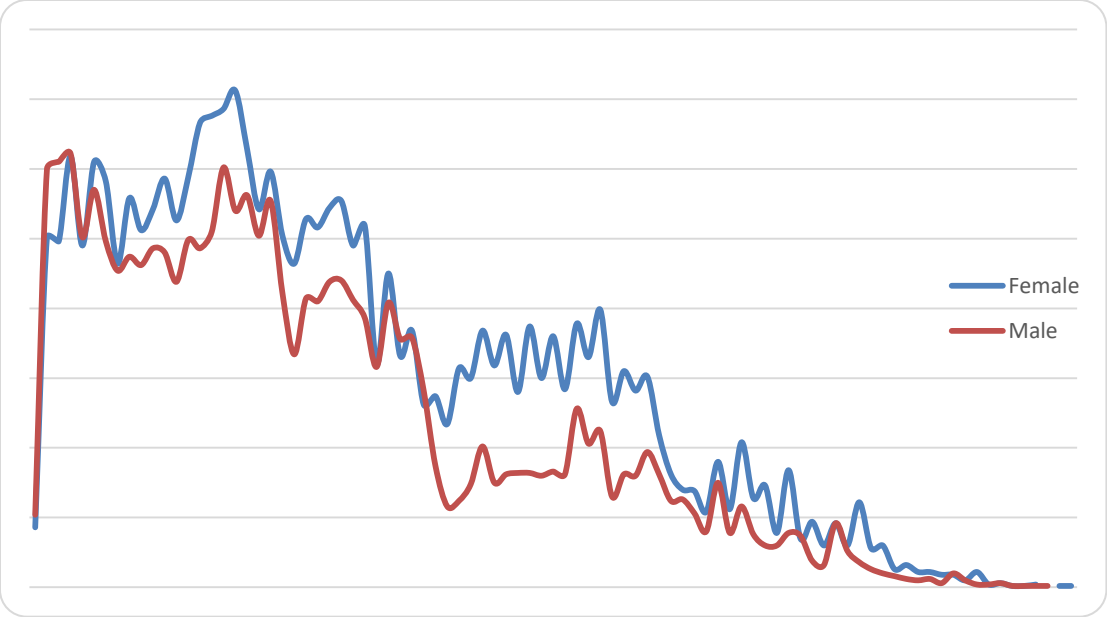

Figure 2

The oldest registrant provided the year of birth as 1908. These registered voters are as seen in Table 6 below. The 2018 FVR for Mutoko North, has 10 (0, 04%) registrants who are above 100 years.

Table 6

Ser	Surname	Forenames	ID Number	Gender	DOB	PS Code
992	KAMUKOSI	ROSA	48-033160 - F48	Female	8/1/1915	4800MTK0103
11043	HOMWE	JANET	48-030681 -L48	Female	7/21/1916	4800MTK0601
2833	CHIWAYA	KABODZI	48-024544 - Q48	Male	5/5/1915	4800MTK0202
5836	BUTE	KASEKE	48-029489 - Q48	Male	5/29/1912	4800MTK0302
13601	HODZI	JULIA	48-037485 - G48	Female	5/15/1910	4800MTK0704
19982	KAVUMBURA	RHODAH	48-008781 - G48	Female	4/7/1915	4800MTK1103
13862	MAPISAUNGA	SIMON	48-029322 -J48	Male	4/1/1917	4800MTK0704
7332	KATIYO	ALICE	49-026499 - X48	Female	10/12/1917	4800MTK0401
14229	USAYI	ZACHARIA	48-004878 - P48	Male	1/21/1916	4800MTK0704
20097	MTIZE	MWANZA	48-001193 -J48	Female	1/12/1908	4800MTK1103

This is an improvement from the 2013 PVR that had 400 persons over the age of 99 as of the end of the year. Only 6 of those 400 records still remain on the roll. Only 2 of the 6 have their ages unaltered from 2013. The rest have had the year of birth changed.

Comparison with the Rolls used in the 30th July 2018 Election

Table 7 gives the breakdown by ward of all registered voters in the current roll:

Table 7

Wards	Male	Female	Grand Total
Ward 1	948	1,043	1,991
Ward 2	1,361	1,576	2,937
Ward 3	1,017	1,268	2,285
Ward 4	767	1,001	1,768
Ward 5	874	1,048	1,922
Ward 6	568	742	1,310
Ward 7	1,010	1,231	2,241
Ward 8	1,013	1,478	2,491
Ward 10	789	1,044	1,833
Ward 11	1,122	1,546	2,668
Ward 12	755	1,094	1,849
Grand Total	**10,224**	**13,071**	**23,295**

Table 8 gives the breakdown of the same in the roll used for the 30th of July 2018 election.

Table 8

Wards	Male	Female	Grand Total
Ward 1	949	1,044	1,993
Ward 2	1,362	1,577	2,939
Ward 3	1,018	1,269	2,287
Ward 4	769	998	1,767
Ward 5	877	1,047	1,924
Ward 6	574	739	1,313
Ward 7	1,013	1,229	2,242
Ward 8	1,017	1,479	2,496
Ward 10	792	1,046	1,838
Ward 11	1,122	1,549	2,671
Ward 12	755	1,093	1,848
Grand Total	**10,248**	**13,070**	**23,318**

While there are changes – these changes are not significant but point to the fact that there is an addition, transfers and update process in place. While there is an aggregate reduction of 23 persons on the current voters roll, the changes are summarized as follows:

- 49 records were removed from the Mutoko North Roll – it's not clear if they were transferred out or were removed completely.
- 26 new registrants were included to the Mutoko North roll of voters. These inclusions were the result of 23 new enrollments and 3 transfers from other constituencies.
- There have been no amendments to voter details (change in names, date of birth, gender etc.) for the rest of the records.

Comparison with the 2013 PVR

When compared with the 2013 PVR for Mutoko North out of the 29,643 persons registered in 2013, only 13,538 registered voters remain in the roll.

The 2013 PVR for Mutoko North had 119 cases of persons sharing both names, dates of birth, gender and polling station. The FVR released before the 2018 election has none but the latest roll has introduced a duplicate record. The case is as seen below.

Surname	Forenames	ID Number	Gender
ZISENGWE	TONNY	48-049846 -S48	Male

Ser	DOB	Voter Address	Polling_Station_Code
7963	7/11/1952	MACHONA VILLAGE, HEADMAN KAWERE, CHIEF MUTOKO, MUTOKO	4800MTK0402
8293	7/11/1957	MACHONA VILLAGE, KAWERE, MUTOKO	4800MTK0403

This seems to be the result of an attempt at amending/update the details of the voter whose ID number is 48-049846 -S48.

Comprehensiveness

Assessment of Over/Underrepresentation

The Age-Gender breakdown of the registered voters is as seen in Table 9 below

Table 9

Age Groups	Male	Female	Grand Total
18 - 19	352	294	646
20 - 24	1,400	1,398	2,798
25 - 29	1,178	1,331	2,509
30 - 34	1,267	1,571	2,838
35 - 39	1,291	1,491	2,782
40 - 44	1,018	1,303	2,321
45 - 49	939	1,055	1,994
50 - 54	527	726	1,253
55 - 59	413	814	1,227
60 - 65	557	1,013	1,570
65+	1,282	2,075	3,357
Grand Total	**10,224**	**13,071**	**23,295**

The Age-Gender breakdown of the projected voting population is as seen in Table 10 below:

Table 10

Age Groups	Male	Female	Grand Total
18 - 19	1,169	964	**2,133**
20 - 24	1,898	2,181	**4,079**
25 - 29	1,813	2,174	**3,987**
30 - 34	1,510	1,638	**3,148**
35 - 39	1,196	1,360	**2,556**
40 - 44	942	942	**1,883**
45 - 49	556	664	**1,219**
50 - 54	473	877	**1,350**
55 - 59	446	734	**1,179**
60 - 65	459	631	**1,089**
65+	1,314	1,641	**2,955**
Grand Total	**11,774**	**13,805**	**25,579**

There are narrow differences between the two tables and Table 11 shows these differences. The figures in red are indicators of gender /age-group combinations where the actual registered

voters exceed the projected registered voters i.e. an over-represented gender/age-group combination. The figures in black indicate gender/age-group combinations where the projected registered voters exceed the actual registered voters - i.e. an under-represented gender/age-group combination.

Table 11

Age Groups	Male	Female
18 - 19	817	670
20 - 24	498	783
25 - 29	635	843
30 - 34	243	67
35 - 39	95	131
40 - 44	76	361
45 - 49	383	391
50 - 54	54	151
55 - 59	33	80
60 - 65	98	382
65+	32	434

An additional viewpoint that compares and contrasts the population pyramids of the eligible voters seen in Figure 3 and registered voters in the constituency seen Figure 4.

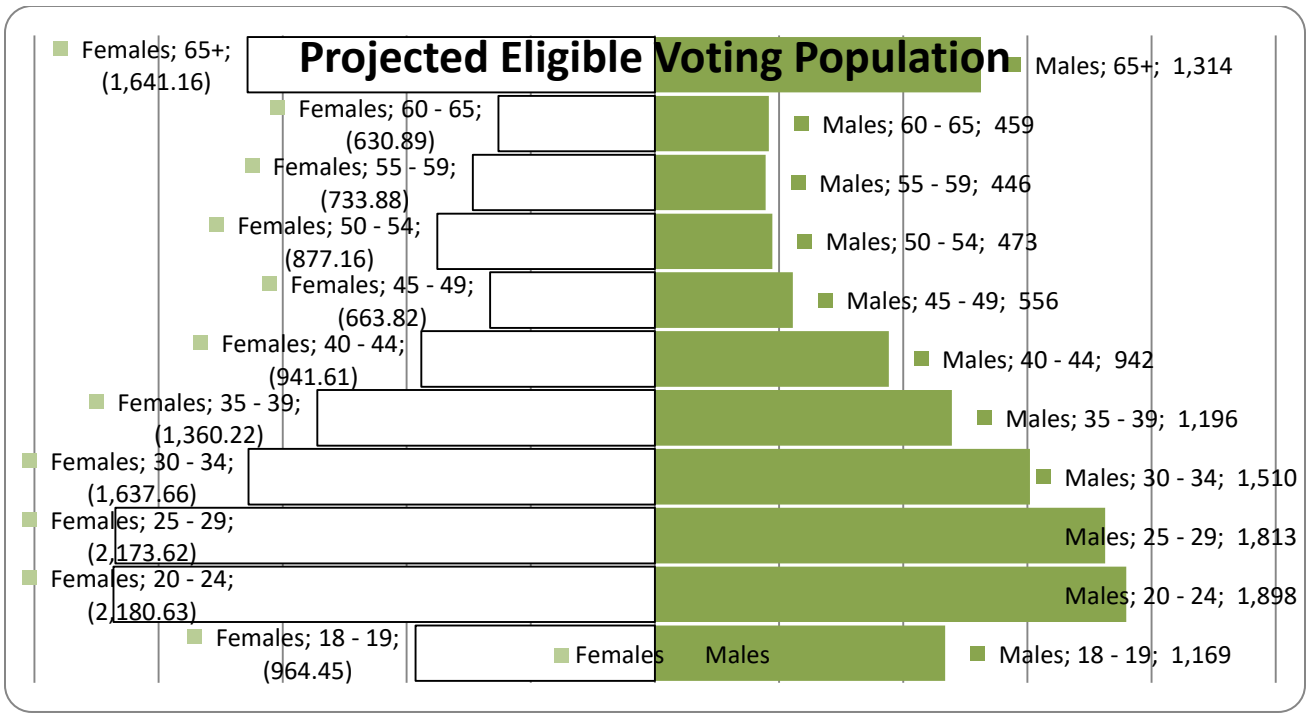

Figure 3

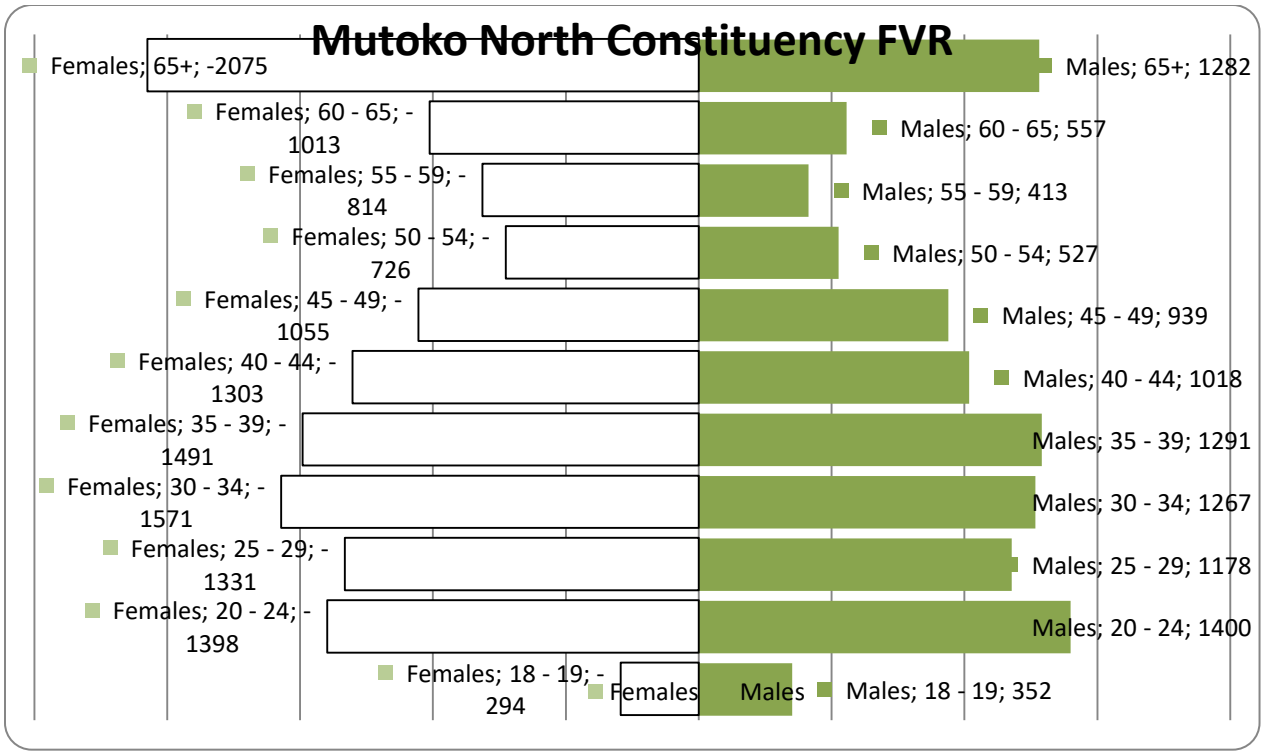

Figure 4

With exception of the 65+ year olds - the projected population layout is more or less a pyramid. This however differs with the registered voters' population "pyramid" seen in Figure 4 below.

ZESN has conducted a projection exercise for voting populations in Zimbabwe and as part of their study there was a ward-by-ward projection. This projection is seen in Table 12 below.

Table 12

Age Group	01	02	03	04	05	06	07	08	10	11	12
18-30	864	1,555	809	1,021	811	339	1,005	590	725	1,039	1,034
31-35	266	427	254	291	251	92	298	158	190	251	305
36-50	428	864	504	563	481	236	622	289	448	528	600
51-98	486	841	640	793	657	301	655	415	570	665	774
N/K	6	14	6	31	4	3	15	5	10	3	38
TOTAL	**2,050**	**3,701**	**2,213**	**2,699**	**2,204**	**971**	**2,595**	**1,457**	**1,943**	**2,486**	**2,751**

Table 13shows the presents the registered voters using the same view. When compared to the ward level analysis using the same parameters.

Table 13

Age Group	01	02	03	04	05	06	07	08	10	11	12
18 - 30	639	917	600	530	521	314	597	593	498	727	499
31 - 35	270	398	277	211	248	160	286	308	242	346	236
36 - 50	555	887	737	450	584	404	695	766	478	753	525
50+	527	735	671	577	569	432	663	824	615	842	589
Total	**1,991**	**2,937**	**2,285**	**1,768**	**1,922**	**1,310**	**2,241**	**2,491**	**1,833**	**2,668**	**1,849**

The figures in red are indicators of ward/age-group combinations where the actual registered voters exceed the projected registered voters. The figures in black indicate ward/age-group combinations where the projected registered voters exceed the actual registered voters.

Age Group	01	02	03	04	05	06	07	08	10	11	12
18 - 30	225	638	209	491	290	25	408	3	227	312	535
31 - 35	4	29	23	80	3	68	12	150	52	95	69
36 - 50	127	23	233	113	103	168	73	477	30	225	75
50+	41	106	31	216	88	131	8	409	45	177	185

It's not significant but there are indeed wards such as 12 and 9 that have under-representation across all the age groups. The same thing happens for the "slight" overrepresentation in ward 8. But these differences are not very significant.

Conclusion

The audit shows that there is a progressive improvement in the quality of the roll from the 2013 PVR to the current roll. The ID number is meant to be unique or at least largely unique i.e. rarely duplicated/difficult to duplicate. The fact that there are duplicated ID numbers in the voters roll points to there being no unique constraint, or at least a mechanism to alert data entry staff of the existence of records sharing this field.

The problem of wrong dates of birth is a rampant problem in Zimbabwe and North Mutoko is no different. These wrong dates of birth make it harder to identify duplicated records in the register.

While it's evident that there is an active process to add, update, remove and transfer voters into the roll – it's not clear when biometric deduplication is used in this case. The case of the duplication of TONNY ZISENGWE (48-049846 -S48) is evidence that the process is not run as often.

The roll is largely representative of the population of Mutoko North since the differences between the actual registered voters and the projected registered voters are not significant.