Quality of Life Survey 6 2020/21

QUALITY OF LIFE SURVEY 6 (2020/21)

SAMPLE DESIGN

NOVEMBER 2021

Authors:
Christian Hamann and Julia de Kadt

GCRO Gauteng City-Region Observatory

GCRO Quality of Life Survey 6 (2020/21): Sample design

Authors: Christian Hamann and Dr Julia de Kadt.

With guidance from: Emeritus Professor Paul Fatti and Professor Mark Orkin

Date: November 2021

Type of output: Technical Report

Research theme: Understanding quality of life

Cover Image: Clive Hassall

Copyright 2021 © Gauteng City-Region Observatory

Published by the Gauteng City-Region Observatory (GCRO), a partnership of the University of Johannesburg, the University of the Witwatersrand, Johannesburg, the Gauteng Provincial Government and organised local government in Gauteng (SALGA).

Suggested citation: Hamann, C. and de Kadt, J. (2021). *GCRO Quality of Life Survey 6 (2020/21): Sample design*. Gauteng City-Region Observatory (GCRO). Johannesburg.

TABLE OF CONTENTS

1 INTRODUCTION

The GCRO Quality of Life (QoL) survey is a biennial survey of randomly selected adults in households across the entire Gauteng Province. The survey is designed to be representative at a ward level[1]. Since the first survey iteration in 2009, the details of the sampling process have varied (Orkin, 2020), but four general stages have remained constant. Stage 1 is the selection of clusters where visiting points should occur, although the nature of clusters have varied over time. Stage 2 is the selection of visiting points in the clusters, stage 3 is the selection of a household at the visiting point and stage 4 is the selection of an adult respondent in the selected household. Stages 1 and 2 are done before going into the field and stages 3 and 4 are done in the field by fieldworkers. This document pertains to the first two stages – the selection of survey clusters within wards and the selection of visiting points. Stages 3 and 4 are documented in the Fieldwork report (GeoSpace International, 2021).

For the Quality of Life Survey 6 (2020/21), random probability proportional to size (PPS) sampling was used to select Enumerator Areas (EAs) within each ward to serve as clusters for visiting points (stage 1). Simple random selection was then used for the selection of residential dwelling units as visiting points (stage 2). This multistage stratified cluster sampling strategy was used as it brings substantial advantages over pure random sampling in terms of the logistical feasibility and cost efficiency of the data collection process (Orkin, 2020). The rest of this document outlines the rationale for the QoL 6 (2020/21) sample design (including key guidelines and the roles of PPS, stratification and substitution), and then presents the detailed design and process to draw the QoL 6 (2020/21) sample.

2 RATIONALE FOR THE QOL 6 (2020/21) SAMPLE DESIGN

2.1 Key parameters

The QoL 6 (2020/21) sample design was guided by learnings from past QoLs, recommendations from the QoL review process (Orkin, 2020) and the available project budget. The key parameters and guidelines for the QoL 6 (2020/21) sample design included the following:

1. Retention of wards as the basis for sampling, both for reasons of comparability with previous QoLs, and due to their ongoing salience to key constituencies.
2. A consistent sample size for wards within each municipality to ensure consistent ward-level precision.
3. A high as possible minimum floor per municipality to ensure adequate precision of estimates in smaller municipalities.
4. A total project budget which allowed for 13 500 interviews.
5. A minimum of 5 clusters per ward.
6. A minimum of 4 visiting points per cluster.

[1] Wards are geopolitical subdivisions of municipalities, delimited by the Municipal Demarcation Board and used for electoral purposes.

A brief overview of the sample distribution implemented on the basis of these guidelines is shown in Table 2.1, and more detailed information is provided in Section 3.3 and Table 3.1.

Table 2.1: A brief overview of the sample design

Municipality	Number of wards	Sample structure	Minimum sample per municipality
City of Johannesburg	135	6 EAs per Ward; a minimum of 4 interviews per EA	3 508
City of Tshwane	107	6 EAs per Ward; a minimum of 4 interviews per EA	2 782
City of Ekurhuleni	112	6 EAs per Ward; a minimum of 4 interviews per EA	2 912
Emfuleni	45	5 EAs per Ward; 4 interviews per EA	900
Lesedi	13	8 EAs per Ward; 6 interviews per EA	624
Merafong	28	5 EAs per Ward; a minimum of 4 interviews per EA	615
Midvaal	15	8 EAs per Ward; 5 interviews per EA	600
Mogale City	39	5 EAs per Ward; 4 interviews per EA	780
Rand West	35	5 EAs per Ward; 4 interviews per EA	700
GAUTENG	529		13 421

2.2 Data sources

The QoL 6 (2020/21) sample design drew on various data sources. The backbone of the sample design is administrative boundaries, of which GCRO holds various spatial layers, including wards (as demarcated in 2016) and EAs (as demarcated and used by Statistics South Africa for the 2011 census). In addition, the sample design drew on data for the location of dwellings units to construct the EA probability variable and to draw visiting points (GeoTerraImage, 2017). More details are provided in the following sections, and a description of the attributes that were used to draw the EA sample is provided in Annexure A (Table 6.1).

2.3 Probability proportional to size (PPS) sampling

Traditionally, PPS uses population data to assign a sampling probability variable to each element of the sample frame. In the QoL 6 (2020/21) sampling process, PPS was only used in the selection of sub-ward clusters (stage 1).

In the QoL 6 (2020/21) sample design, the total number of dwelling units was used to determine the probability variable. Since we sampled dwelling units, it is appropriate that the PPS selection of areas was also related to dwelling units. The probability variable in the QoL 6 (2020/21) sample design was the proportion of all dwelling units in the ward that are located in the particular sub-ward area. In other words, sub-ward areas with a higher proportion of the total dwelling units in the ward had a higher probability of being selected (although selection is not guaranteed).

2.4 Stratification

Stratified random sampling is efficient when the sample population can be easily broken into distinct sub-groups, and it has been used in previous QoL surveys. Samples are then taken from each sub-group based on the ratio of the sub-group's size to the total population. This provides higher statistical precision than simple random sampling, prevents the exclusion of any particular sub-group and ensures that the characteristics of each sub-group are included in the sample. However, stratification may have unknown consequences because it requires one to assign a generalised attribute to areas in order to assign them to strata, before sampling the appropriate number of areas within each strata. In the QoL review, Professor Mark Orkin noted, "in principle, with a clear and salient stratifying variable to hand, stratifying seems an obvious choice. In practice, given possible [data] anomalies and especially given the extra effort..." it is doubtful whether stratification is worth it (Orkin, 2020: 24). In other words, EAs are unlikely to be as homogeneous as a generalised attribute suggests and it is likely that members of secondary groups will be interviewed.

GCRO was in the possession of a demographics data set, acquired from GeoTerraImage, which had the potential of assisting with sub-ward stratification. "The demographics data set provides insight into the current demographic distribution per EA in terms of Gender, Race and Age Group as well as the dominant group in each of these categories" (GeoTerraImage, 2020a) and could therefore have been used to stratify the selection of EAs by population group. During the selection of EAs, GCRO used the abovementioned data to test the effect that stratification might have on the sample if stratification by population group was implemented. According to Figure 2.1, on a provincial level, the initial EA sample (before any in-field substitution took palce) over-sampled African EAs by 14 and Indian EAs by 7, while under-sampling Coloured EAs by 7 and White EAs by 14. This is considered insignificant in a total sample of 3 075 EAs. Within any ward in Gauteng the over- or under-sampling of any dominant population group did not exceeded 2 EAs per ward, except in one ward where the difference was 3 EAs. Additionally, this level of over- or under-representation in the final survey sample, after fieldwork, would be corrected through the weighting process (Neethling, 2021). Further, given some data anomalies in the demographics data set, the added effort of implementing stratification and the insignificant variation between the non-stratified EA sample and probable stratification-based EA sample, it was decided that stratification was not required and therefore not pursued further.

Figure 2.1: Comparing the results of possible stratification with the results after sampling without stratification

Municipality	Number of Eas in each municipality, by dominant population group in each EA					EAs to be sampled if stratification is implemented				EAs sampled without stratification				Difference between no stratification and stratification			
Dominant population group of the EA	African	Coloured	Indian	White	Total	African	Coloured	Indian	White	African	Coloured	Indian	White	African	Coloured	Indian	White
Ekurhuleni	3696	67	45	1041	4849	535	10	6	121	530	10	6	126	-5	0	0	5
Johannesburg	4485	223	199	1229	6136	637	30	25	115	636	26	28	117	-1	-4	3	2
Tshwane	3465	35	35	1277	4812	474	5	4	159	483	3	8	148	9	-2	4	-11
Emfuleni	1015	3	8	173	1199	196	1	1	27	200	1	1	23	4	0	0	-4
Lesedi	156	0	1	42	199	83	0	0	21	87	0	0	17	4	0	0	-4
Midvaal	142	0	0	65	207	76	0	0	41	76	0	0	41	0	0	0	0
Merafong	330	0	0	31	361	125	0	0	13	125	0	0	13	0	0	0	0
Mogale City	460	0	9	158	627	147	0	4	44	151	0	4	40	4	0	0	-4
Rand West	478	24	0	90	592	139	11	0	25	138	10	0	27	-1	-1	0	2
GAUTENG	14227	352	297	4106	18982	2412	57	40	566	2426	50	47	552	14	-7	7	-14

2.5 Substitution

Clusters (or EAs) and primary visiting points were drawn before fieldwork commenced. However, given the access challenges and response rates usually encountered in Gauteng, it was always anticipated that substitutions and secondary visiting points would be required in order to complete the desired sample. If an entire EA proved to be inaccessible (for example when access to a security estate or farm was not granted) the substitution was done on a case by case basis. The inaccessible EA was substituted with another EA that was drawn on the same principles described in section 3.3, below.

The substitution of visiting points occurred when a particular primary visiting point proves inaccessible. For each primary visiting point, three substitution visiting points were drawn before fieldwork commenced and were provided to the service provider. Substitution visiting points were also drawn from the data for the location of dwellings units (GeoTerraImage, 2017), after selection of the primary visiting points and their exclusion from the dataset. Substitution visiting points were drawn on the same principles described in section 3.5, below. A high-level overview of the extent of substitutions at the EA and visiting point levels is provided in the QoL 6 Fieldwork Report (GeoSpace International, 2021).

3 THE QOL 6 (2020/21) SAMPLE DESIGN

3.1 Sub-ward geography

The ideal sub-ward geography for QoL is one that is drawn around residential settlements, does not include non-contiguous polygons, covers the entire Gauteng province, is relatively small in geographic extent, is closely aligned to ward boundaries, and can be related to official population data (like data released by Statistics South Africa). All of these characteristics are not currently available in a single sub-ward geography and therefore the GCRO had to weigh the advantages and disadvantages of various sub-ward geographies before settling on the one most suited to QoL 6 (2020/21).

There were various sub-ward geographies to choose from. Initial options included sub-place boundaries, a Small Area Layer (SAL, also used in the 2011 census), EAs, Voting District boundaries and the creation of a custom geography. Sub-place boundaries were ruled out because they vary significantly in size and because the layer includes non-contiguous polygons. The creation of a custom sub-ward geography was not pursued because it was deemed as an unnecessary effort when other potentially viable geographies were available. Therefore, consideration was focused on the SAL, EA and Voting District geographies (Table 3.1).

After careful consideration, in consultation with GCRO's QoL 6 (2020/21) external advisers as well as the appointed service provider, EAs were chosen as the most suitable sub-ward geography for QoL 6 (2020/21). EA boundaries are drawn around residential areas, are the smallest available sub-ward geography, and cover the entire Gauteng province. SAL boundaries are also very small, but do not cover the entire Gauteng Province and were therefore not selected. Voting districts have the benefit of being exactly aligned to ward boundaries but their average size is much too large for the purposes of the QoL 6 (2020/21) sampling strategy.

Table 3.1: Sub-ward geographies considered for QoL 6 (2020/21)

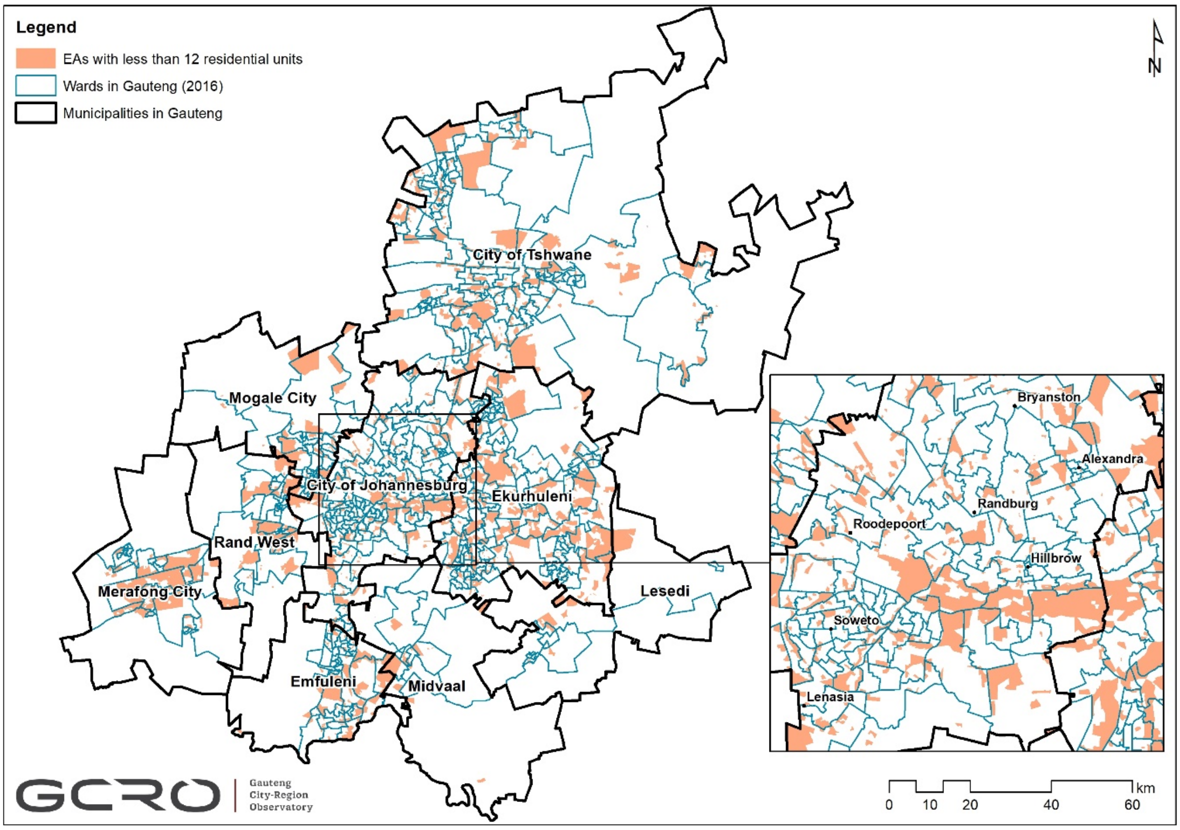
Sub-ward geography	Drawn specific to residential developments	Includes non-contiguous polygons	Covers entire Gauteng Province	Average size (km²)	Aligned with wards	Related to official population data
EA	Yes	No	Yes	0.8715	No	No
SAL	Yes	No	No	0.9720	No	Yes
Voting district	Yes	No	Yes	6.5573	Yes	Yes

3.2 The EA sample frame

The first stage of the sampling process was to draw the sub-ward geographies in which interviews were clustered. The starting point was all the EAs in Gauteng (n = 20 850). All the EAs with less than 12 residential dwelling points [2], as per the data for the location of dwellings units (GeoTerraImage, 2017) were excluded from the sample frame (n = 1 867, Figure 3.1). This included some EAs which are sparsely populated, but mainly included EAs drawn around nature reserves, mining areas, industrial parks and retail or school precincts. The random PPS sample of EAs was drawn from the remaining **18 983** EAs, using the attributes detailed in Annexure A (Table 6.1).

[2] If we draw 4 primary visiting points per EA and 3 potential replacement points, we require a minimum of 12 potential visiting points. It is too costly to risk going to EAs that do not have enough potential visiting points.

Figure 3.1: EAs with less than 12 residential dwelling points and therefore not eligible for the sample



3.3 Drawing the EA sample

Using this EA sample frame, the required number of EAs per ward were randomly selected for each ward in Gauteng. This process was undertaken using Microsoft Excel. Key steps in the selection process were:

a. The first calculation was to determine, for each EA, the proportion of the total ward dwelling units represented. EAs were then sorted from low to high according to the dwelling unit proportions within each ward.
b. The cumulative proportion of dwelling units per ward was calculated.
c. In order to get the desired result with the excel formula used for the random selection, EAs were re-sorted from high to low cumulative values within each ward.
d. Random numbers were generated in the sample frame and random numbers were matched with the PPS variable and returned a randomly selected EA (Figure 3.2). A single formula in excel was used for this selection.
e. The drawn EAs were marked in the final sample for mapping (Figure 3.3).

Figure 3.2: An example of drawing random EAs within a ward with the sample frame and formula

EA_CODE	Centroid_ward	Total_EA_units	Ward_Units	P_Units_2	Cumulative	RANDOM	EA_CODE_copy	INDEX_MATCH	Sampled_EAs	Dominant_group_sampled	MN_NAME	DC_NAM
76011114	74201024	349	3615	0.09654219	1.00000000	0.52562569	76011114		=VLOOKUP(INDEX($G$694:$G$709,MATCH(H694,$G$694:$G$709,-1),1),			
76011112	74201024	297	3615	0.08215768	0.90345781	0.17572639	76011112		$G$694:$I$709,3,FALSE)			
76011113	74201024	292	3615	0.08077455	0.82130014	0.25931324	76011113		76011101	BLACK	Emfuleni	Sediben
76011106	74201024	239	3615	0.06611342	0.74052559	0.75429663	76011106		76011113	BLACK	Emfuleni	Sediben
76011100	74201024	232	3615	0.06417704	0.67441217	0.31741312	76011100		76011096	BLACK	Emfuleni	Sediben
76011103	74201024	231	3615	0.06390041	0.61023513	0.32804262	76011103			0	Emfuleni	Sediben
76011348	74201024	231	3615	0.06390041	0.54633472	0.93056129	76011348			0	Emfuleni	Sediben
76011104	74201024	229	3615	0.06334716	0.48243430	0.16198845	76011104			0	Emfuleni	Sediben
76011108	74201024	219	3615	0.06058091	0.41908714	0.49335972	76011108			0	Emfuleni	Sediben
76011096	74201024	217	3615	0.06002766	0.35850622	0.72667578	76011096			0	Emfuleni	Sediben
76011101	74201024	202	3615	0.05587828	0.29847856	0.68968160	76011101			0	Emfuleni	Sediben
76011105	74201024	196	3615	0.05421853	0.24260028	0.67514306	76011105			0	Emfuleni	Sediben
76011111	74201024	187	3615	0.05172891	0.18838174	0.44836392	76011111			0	Emfuleni	Sediben
76011109	74201024	178	3615	0.04923928	0.13665284	0.34658496	76011109			0	Emfuleni	Sediben
76011110	74201024	160	3615	0.04426003	0.08741355	0.04402765	76011110			0	Emfuleni	Sediben
76011097	74201024	156	3615	0.04315353	0.04315353	0.63471003	76011097			0	Emfuleni	Sediben

Figure 3.3: Identifying the final sample for mapping

EA_CODE	Sample	Centroid_ward	Total_EA_units	Ward_Units	P_Units_2	Cumulative	RANDOM	EA_CODE_copy	INDEX_MATCH	Sampled_EAs	Dominant_group_sampled	EA_Selected	MN_
76011114	0	74201024	349	3615	0.09654219	1.00000000	0.52562569	76011114		76011348	BLACK	1	Emfu
76011112	0	74201024	297	3615	0.08215768	0.90345781	0.17572639	76011112		76011111	BLACK	1	Emfu
76011113	=IFERROR(VLOOKUP(B696,$L$2:$N$18986,3,FALSE),0)				0.08077455	0.82130014	0.25931324	76011113		76011101	BLACK	1	Emfu
76011106	0	74201024	239	3615	0.06611342	0.74052559	0.75429663	76011106		76011113	BLACK	1	Emfu
76011100	0	74201024	232	3615	0.06417704	0.67441217	0.31741312	76011100		76011096	BLACK	1	Emfu
76011103	0	74201024	231	3615	0.06390041	0.61023513	0.32804262	76011103			0	0	Emfu
76011348	1	74201024	231	3615	0.06390041	0.54633472	0.93056129	76011348			0	0	Emfu
76011104	0	74201024	229	3615	0.06334716	0.48243430	0.16198845	76011104			0	0	Emfu
76011108	0	74201024	219	3615	0.06058091	0.41908714	0.49335972	76011108			0	0	Emfu
76011096	1	74201024	217	3615	0.06002766	0.35850622	0.72667578	76011096			0	0	Emfu
76011101	1	74201024	202	3615	0.05587828	0.29847856	0.68968160	76011101			0	0	Emfu
76011105	0	74201024	196	3615	0.05421853	0.24260028	0.67514306	76011105			0	0	Emfu
76011111	1	74201024	187	3615	0.05172891	0.18838174	0.44836392	76011111			0	0	Emfu
76011109	0	74201024	178	3615	0.04923928	0.13665284	0.34658496	76011109			0	0	Emfu
76011110	0	74201024	160	3615	0.04426003	0.08741355	0.04402765	76011110			0	0	Emfu
76011097	0	74201024	156	3615	0.04315353	0.04315353	0.63471003	76011097			0	0	Emfu

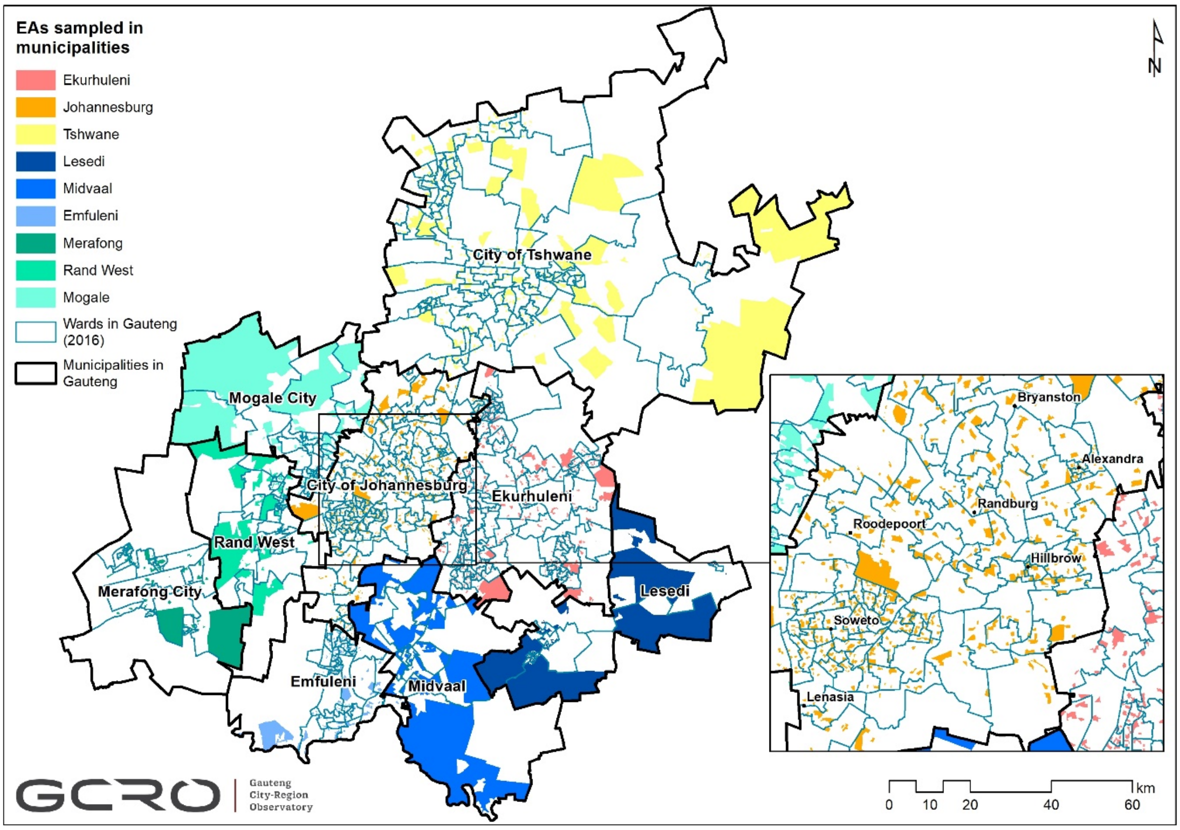
Within a ward, instances where duplicate EAs were drawn were immediately identified. In these cases, a different random number was used until a set of unique EAs were drawn. A total of 3 075 EAs were selected to form the clusters within which visiting points were drawn. The number of EAs sampled in each ward per municipality is described in more detail in Table 3.2 and the spatial distribution is shown in Figure 3.4.

The selected EAs were mapped, but in some cases their boundaries did not coincide with ward boundaries. In order to keep visiting points within the correct ward, the EA boundaries were clipped to the ward where the EA centroid is located. This effectively reduced the "available size" of the selected EA and therefore a further check as done to ensure that all the selected EAs had enough potential visiting points. A total of 12 EAs no longer had enough potential visiting points. These were removed from the sample, and replaced using the same random selection process described above.

Table 2.2: The sample distribution of QoL 6 (2020/21)

Municipality	Number of wards	Total number of EAs	Number of sampled EAs	Sample structure	Visiting points per ward	Visiting points per municipality	Implementation notes
City of Johannesburg	135	6 135	807	6 EAs per Ward; 4 interviews per EA	26	3 508	Increased from 5 to 6 EAs per ward. Drew 8 visiting points per EA in one ward (ward: 79800059). Drew 1 additional visiting point in every 3rd EA in other wards.
City of Tshwane	107	4 812	642	6 EAs per Ward; 4 interviews per EA	26	2 782	Increased from 5 to 6 EAs per ward. Drew 1 additional visiting point in every 3rd EA.
City of Ekurhuleni	112	4 849	672	6 EAs per Ward; 4 interviews per EA	26	2 912	Increased from 5 to 6 EAs per ward. Drew 1 additional visiting point in every 3rd EA.
Emfuleni	45	1 200	225	5 EAs per Ward; 4 interviews per EA	20	900	
Lesedi	13	199	104	8 EAs per Ward; 6 interviews per EA	48	624	Drew 3 additional EAs per ward and 2 additional interview per EA.
Merafong	28	362	138	5 EAs per Ward; 4 interviews per EA	22 (one with 21)	615	Drew 7 visiting points per EA in one ward (ward: 74804019). Drew 1 additional visiting point in every 2nd and 4th EA in other wards.
Midvaal	15	207	117	8 EAs per Ward; 5 interviews per EA	40	600	Drew 3 additional EAs in 14 wards because there are only 14 wards where this is possible and drew 3 additional visiting points per EA in the 15th ward.
Mogale City	39	627	195	5 EAs per Ward; 4 interviews per EA	20	780	
Rand West	35	591	175	5 EAs per Ward; 4 interviews per EA	20	700	
GAUTENG	529	18 982	3 075			13 421	There were still 79 visiting points against the target of 13 500. These were used flexibly across the province when implementation varied slightly from plans.

Figure 3.4: The random selection of Enumerator Areas (EAs) for the QoL 6 (2020/21) sample

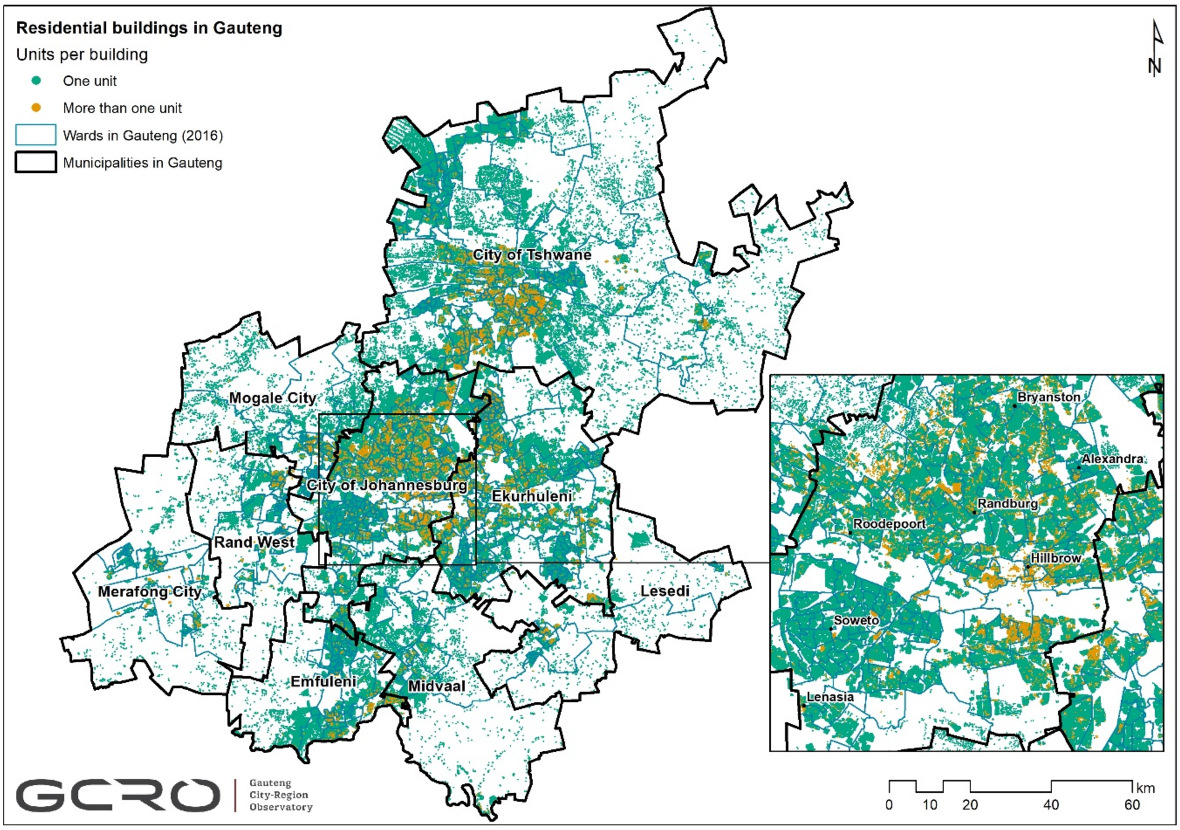


3.4 The dwelling unit sample frame

Following the selection of the EAs comprising the sample, the next step was the selection of visiting points in each sampled EA. Visiting points were drawn from the GeoTerraImage Building Based Land Use 2017 dataset. GeoTerraImage provided the data as a building based land use point data set in which "the land use is classified into 70 land use classes, identifying every structure [in the Gauteng Province] according to a set of comprehensive land use definitions" (GeoTerraImage, 2020b).

The starting point was to limit this dataset to only include residential buildings in Gauteng (n = 3 394 110). Children's homes and correctional facilities were then excluded (n = 73). The remaining buildings (n = 3 394 037) were duplicated by the number of dwelling units associated to each building. In other words, if a block of flats or cluster housing complex had 50 units, the point was duplicated 50 times to represent the entire universe of available dwelling units. The final list of dwelling units from which the sample was drawn included **4 050 379** dwelling units (Figure 3.5).

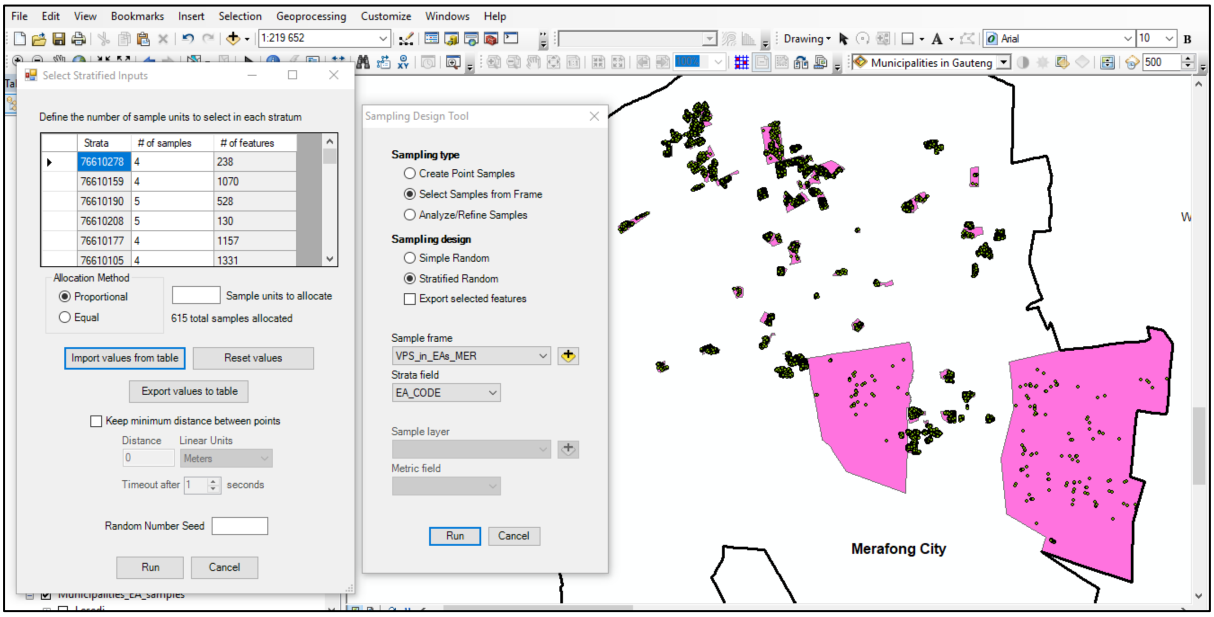
Figure 3.5: The distribution of dwelling units in Gauteng (Source: GeoTerraImage, 2017)



3.5 Drawing the dwelling units for visiting points

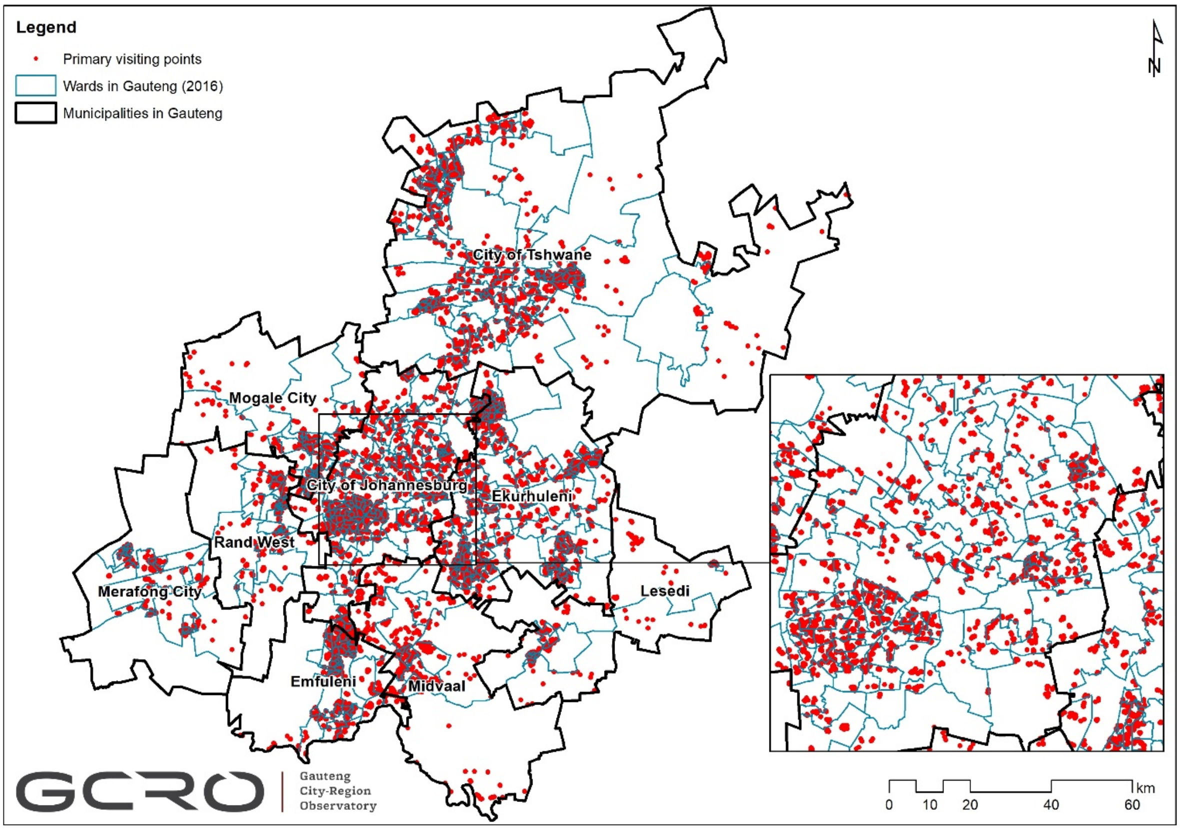
Stage 2 of sample selection involved simple random selection of visiting points from the dwelling unit sample frame described above. The random selection was handled by the *Sampling Design Tool for ArcGIS* (Buja and Menza, 2013; Figure 3.6) and done for individual municipalities at a time. The *Sampling Design Tool for ArcGIS* randomly selects visiting points from the selected sampling frame. EAs (selected in stage 1) were used as the strata and the required number of visiting points (outlined in Table 3.2) was assigned to each EA. The tool then randomly selected visiting points from the available dwelling units.

Figure 3.6: Using the *Sampling Design Tool for ArcGIS* to draw random visiting points in Merafong City



The final sample of primary visiting points (n = 13 421) is shown in Figure 3.7. The selected primary visiting points were removed from the sample frame and the remaining dwelling units became eligible for selection as substitution points. Substitution points within each EA were drawn in the same way as the primary visiting points.

Figure 3.7: The random selection of primary visiting points for the QoL 6 (2020/21) sample

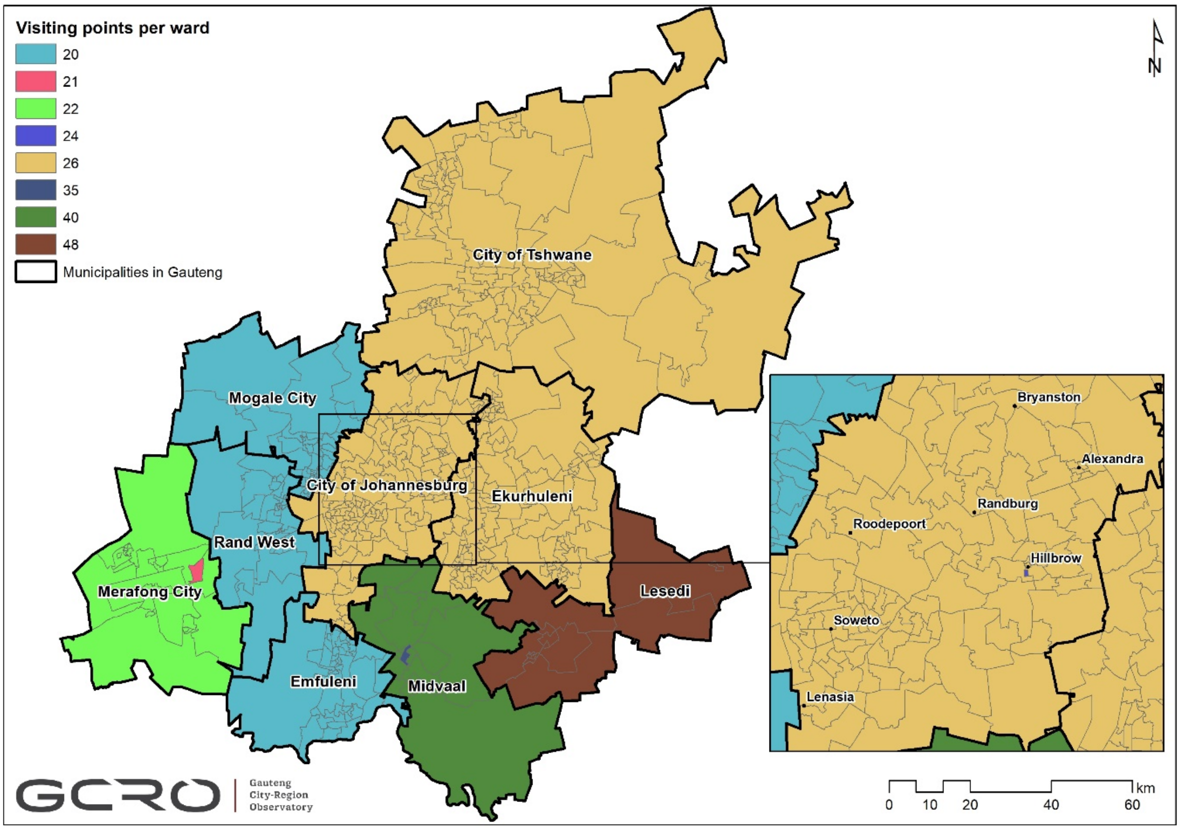


The number of primary visiting points per ward and municipality is shown in Figure 3.8[3]. In some instances, selected wards within some municipalities have a lower number of primary visiting points than other wards in the same municipality. This is the result of various efforts described in the notes column in Table 3.2 to balance the sample and reach the minimum required visiting points for each municipality. Additional differences between wards within a municipality occurred in two municipalities:

- In Merafong City, we drew 7 visiting points per EA in one ward (WARD ID: 74804019) because there were only 3 valid EAs to sample from in the ward. This amounted to 21 primary visiting points in the ward as opposed to the 22 primary visiting points in the other of the wards in the municipality.
- In the City of Johannesburg, we drew 8 visiting points per EA in one ward (WARD ID: 79800059) because there were only 3 valid EAs to sample from in the ward. The ward is located in central Johannesburg and includes only high-rise apartment buildings. This resulted in 24 primary visiting points in the ward as opposed to the 26 primary visiting points in the rest of the wards in the municipality.

[3] Note that the final sample that was attained is slightly different from this sampled distribution, due to the nature of fieldwork. Additionally, in a number of wards, slightly more interviews than planned were completed.

Figure 3.8: The number of primary visiting points per ward for the QoL 6 (2020/21) sample



3.6 Drawing the substitution points

For each primary visiting point, four substitution points were drawn in the same EA. The sample frame for the substitution points excluded multi-unit buildings which were already sampled. This prevented substitution points from being inside a building where access had already been denied, and also meant field teams did not need to negotiate access to a building at more than one time.

Instead, once access to a multi-unit building was negotiated, if refusals were encountered in the building, teams would identify substitute points in the same building. In field substitution across multi-unit buildings also took place, but only after all options in the building that was first selected were exhausted. At that point, the "substitution buildings" was used to gather the remaining required interviews in the EA.

4 CHANGES DURING FIELDWORK

During fieldwork, it was necessary to adapt to unplanned challenges and delays experienced. A high rate of refusals in high wall areas and hostels proved to be problematic and slowed down fieldwork progress. In many cases in these areas, sampled primary and substitute visiting points were exhausted through outcomes such as access refusals or 'no one at home'. Initially, additional substitution points were provided as needed, but this tended to delay fieldwork progress in the relevant EA. In some instances, it even meant repeating gate keeper meetings in order to work in the same area. Therefore, adaptations were made to improve fieldwork efficiency in particularly challenging EAs.

In these areas, all residential points were made available to the service provider, and loaded into the fieldworker management system by the service provider. A random number was assigned to each point which was used as a priority ranking. In each of the incomplete EAs, the starting point was the point with the lowest random value. The fieldworker moved to the next lowest random value for the next interview until all the required interviews has been completed for the EA. The random ranking numbers was visible on the mobile maps used by fieldworkers for easy navigation purposes.

5 REFERENCES

Buja, K., and Menza, C., 2013. *Sampling Design Tool for ArcGIS: Instruction Manual*. Available online at: https://www.arcgis.com/home/item.html?id=f7289cfc69204aa688e8c7c739fc0901 (Accessed 11/05/2020).

GeoSpace International, 2021. *GCRO Quality of Life Survey 6: Fieldwork report (2020/21)*. Johannesburg: Gauteng City-Region Observatory. Available online at: https://gcro.ac.za/research/project/detail/quality-life-survey-vi-202021/

GeoTerraImage, 2017. *Building Based Land Use point data set*. Spatial layer, points represent buildings.

GeoTerraImage, 2018. *Demographics estimates: EA summary population*. Spatial layer, polygons are Statistics South Africa 2011 EA boundaries.

GeoTerraImage, 2020a. *Demographics estimates*. Available online at: http://www.geoterraimage.com/products-demographics.php (Accessed 07/05/2020).

GeoTerraImage, 2020b. *Land Use Data Products: Building Base Land Use*. Available online at: http://www.geoterraimage.com/products-landuse.php (Accessed 07/05/2020).

Neethling, A., 2021. *GCRO Quality of Life Survey 6: Weighting report (2020/21)*. Johannesburg: Gauteng City-Region Observatory. Available online at: https://gcro.ac.za/research/project/detail/quality-life-survey-vi-202021/

Orkin, M., 2020. *Technical Review of the GCRO QoL Surveys: Synthesis Report*. Johannesburg: Gauteng City-Region Observatory. Available online at: https://www.gcro.ac.za/research/project/detail/quality-life-survey-10-year-review/

6 ANNEXURE A

Table 6.1: A description of attributes used to draw the EA sample

Variable	Description	Notes
EA_CODE	EA code (2011)	Source: Statistics South Africa
Centroid_ward	The WARD_ID where the centroid of the EA is located	Calculated
Total_EA_units	Sum of dwelling units per EA	Source: GeoTerraImage, 2017
Ward_Units	Sum of dwelling units per Ward	Source: GeoTerraImage, 2017
P_Units_2	Proportion of Ward_Units located in each EA	Calculated
Cumulative	Cumulative proportion of dwelling units in the ward	Calculated
RANDOM	Random number generated in Microsoft Excel	Calculated
EA_CODE_copy	A copy of EA_CODE	Used to simplify the implementation of the random selection formula
Sampled_EAs	Formula used for the selection of EAs for the sample (as depicted in Figure 3)	Calculated
Dominant_group_sampled	Primary dominant population group of the sampled EA	Source: GeoTerraImage, 2018
MN_NAME	Municipality name	
DC_NAME	District name	
D_P_RACE	Primary dominant population group per EA	Source: GeoTerraImage, 2018