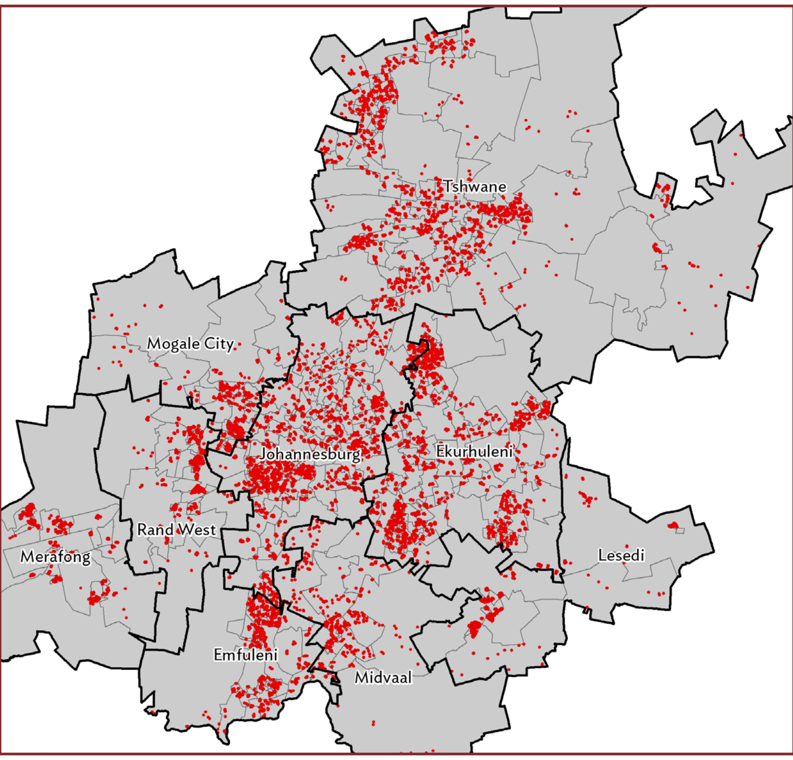

Quality of Life Survey 6 2020/21

QUALITY OF LIFE SURVEY 6 (2020/21)

WEIGHTING REPORT

NOVEMBER 2021

Author:
Ariane Neethling

Gauteng City-Region Observatory

Quality of Life Survey 6 (2020/21): Weighting report

Authors: Ariane Neethling[1]

Date: November 2021

Type of output: Technical Report

Research theme: Understanding quality of life

Cover Image: Christian Hamann

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: Neethling, A. (2021). *Quality of Life Survey 6 (2020/21): Weighting report*. Gauteng City-Region Observatory (GCRO). Johannesburg.

[1] Statistical consultant and part-time lecturer, Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa.
Email: ariane_neethling@yahoo.com

TABLE OF CONTENTS

1 INTRODUCTION

Weights are assigned to make weighted sample records represent the target population as accurately as possible. A weight (w_i) indicates the number of population elements "represented" by a single sample element. Therefore, the sum of the weights $\sum w_i = N$ should be equal to the population total.

Weights are usually developed in different stages to compensate for:

- Unequal inclusion probabilities due to the design, by calculating the design weights.
- Non-response, by adjusting the design weight, if necessary.
- Non-coverage and skewness resulting from, inter alia, fieldwork, by using for example cell-weighting, rim weighting or other calibration techniques.

This document provides technical detail on the processes followed to calculate weights for the GCRO Quality of Life Survey 6 (2020/21). Weights were calculated using adult population and household count estimates provided by GeoTerraImage (GTI).

Weights should be used during analysis of the QoL 6 (2020/21) data as detailed in *Quality of Life Survey 6 (2020/21): Analyses under complex sampling* (Neethling, 2021).

2 WEIGHT VARIABLES FOR THE QOL 6 SURVEY

Two sets of weights are included in the QoL 6 (2020/21) survey dataset – an individual weight for use when results are required per individual, and a household weight when conclusions are required in terms of households. For example, to determine the estimated percentage of *households* without running water inside the house, use the household weight. The use of the individual weight is required for the estimated percentage of *people* without running water inside the house. It is essential to use the weight variable which best suits your desired purposes, because results will vary, even when presented in percentage form, depending on which weight variable is used.

The weighting variable 'DOWNSCALE_MUN_PP_BENCHWGT' provides an individual level weight, scaled to the QoL 6 (2020/21) sample size of 13 616, after it was calculated to sum to the population total. This is the default variable used in GCRO analysis, and is appropriate for use in all individual level analyses. Note that this weight can only be used for percentage (proportions) and mean estimates. If an estimate of actual population size is desired, the original weight (before it was downscaled) should be used. The weighting variable 'HH_WEIGHT' provides a household level weight. This weight is suitable for use in any household level analyses. This weight has not been downscaled, so the frequency figures are the estimated total number of households in Gauteng province.

Further information of the correct application of these weights for data analysis is provided in *Quality of Life Survey 6 (2020/21): Guide to weighted data analysis* (Neethling, 2021). Queries can also be directed to the GCRO.

3 DESIGN WEIGHTS

In order to obtain a representative sample of the population, a stratified multistage probability sample was designed with ward boundaries as the stratification variable and Enumerator Areas (EAs) as the primary sampling units. More details about the sample design are provided in the sample design report (Hamann &

de Kadt, 2021). Since the sample was designed within each stratum (ward) in three stages, the design weight of a household and respondent should be calculated according to the inclusion probability of a unit at each stage. This has to be done within a stratum (Cochran (1977); Lohr (2010)) .

First stage

In the first stage of sampling, primary sampling units (psu), i.e. the EAs, were selected with probability proportional to size (pps) from the population sampling frame. Thus, the weight of an EA, i.e. the inverse of the inclusion probability of an EA, is given by

$$w_{EA} = \left(n_{PSU} \frac{EA_{MOS}}{POP_{MOS}}\right)^{-1}$$

Where:

n_{PSU} is the number of EAs selected in the stratum,

EA_{MOS} is the measure of size (MOS), thus the number of dwelling units in a selected EA, and

POP_{MOS} is the total MOS, thus the total number of dwelling units, of all the EAs in the stratum.

Second stage

The second stage of sampling was the selection of a predetermined number of households, with equal probability, from each selected EA. The household weight per EAis given by

$$w_{HH} = w_{EA} \left(\frac{n_{HH}}{EA_{HH}}\right)^{-1},$$

Where:

n_{HH} is the number of selected households per EA, and

EA_{HH} is the total number of households in the selected EA.

Third stage

In the third stage of sampling (this occurs in the field), a person aged 18 years or older was randomly selected from a household to be interviewed. The respondent weight was calculated by

$$w_{PP} = w_{HH} \times n_{18+}^{*},$$

Where:

n_{18+}^{*} is the average number of persons aged 18 years and older in households in the EA.

The average household size for the EA is used, instead of the observed number of persons 18 years and older in any particular household, to obtain smoother design weights.

4 CALIBRATION

The design weights of the respondents (as calculated above) were adjusted to compensate for differential non-response (i.e. under/over-representation of certain parts of the population). Calibration estimation has become a widely used method for obtaining efficient estimates in sampling surveys by using auxiliary information in the form of known population totals to produce a new set of weights, called calibration weights. For more information about different calibration methods, their formulae and characteristics, see Deville and Särndal (1992); Deville et al. (1993); Neethling (2004) and Neethling & Galpin (2006). Commonly used methods, such as cell weighting, poststratification and rim weighting are special cases of calibration. The SAS macro CALMAR, developed by INSEE in France, was used to adjust the design weights to the newest released GeoTerraImage (GTI) 2021 population estimates, which are based on the 2020 mid-year estimates of Statistics South Africa.

The calibration technique is applied for each local or metropolitan municipality in Gauteng, with race-by-gender and wards as benchmark variables. This means that the final set of weights sums to the correct population totals per ward as well as race-by-gender numbers in each local municipality. Due to too few Indian/Asian respondents in the local municipalities of Emfuleni, Midvaal, Lesedi, West Rand, the Indian/Asian males and females for all six local municipalities sum to the figure for the relevant district municipality. For Merafong, Coloured males and females were also combined due to small sample sizes. Table 1.1 summarises the race-by-gender groups used in the benchmarking phase.

Table 1.1: Race-by-gender groups used in the benchmarking phase

<table>
<tr><th rowspan="2">Municipality</th><th colspan="8">Population group</th></tr>
<tr><th colspan="2">Black African</th><th colspan="2">Coloured</th><th colspan="2">Indian/Asian</th><th colspan="2">White</th></tr>
<tr><td>Emfuleni</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td rowspan="3">Male</td><td rowspan="3">Female</td><td>Male</td><td>Female</td></tr>
<tr><td>Midvaal</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td></tr>
<tr><td>Lesedi</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td></tr>
<tr><td>Mogale City</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td rowspan="3">Male</td><td rowspan="3">Female</td><td>Male</td><td>Female</td></tr>
<tr><td>Merafong</td><td>Male</td><td>Female</td><td colspan="2">Both</td><td>Male</td><td>Female</td></tr>
<tr><td>Rand West</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td></tr>
<tr><td>Ekurhuleni</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td></tr>
<tr><td>Johannesburg</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td></tr>
<tr><td>Tshwane</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td><td>Male</td><td>Female</td></tr>
</table>

Since the Statistics South Africa mid-year estimates do not include the population group 'Other', the GTI database also does not include 'Other'. In the GCRO QoL 6 sample, only 17 of the 13 616 respondents were captured in the 'Other' population group. For this reason, the dominant population group of the EA in which they were interviewed, was assigned to these individuals for the benchmarking phase. The results for unweighted and weighted percentages of the sample per municipality are provided in Table 1.2 and similar data for race and sex groups are provided in Table 1.3.

Table 1.2: Municipal sample distribution for unweighted and weighted QoL 6 (2020/21) data

Municipality	Sample size (unweighted)	Percentage of sample (unweighted)	Percentage of sample (weighted)
City of Johannesburg	3 545	26.0%	38.4%
City of Tshwane	2 810	20.6%	24.1%
City of Ekurhuleni	2 963	21.8%	25.4%
Emfuleni	907	6.7%	4.6%
Lesedi	647	4.8%	0.7%
Merafong	631	4.6%	1.6%
Midvaal	606	4.5%	0.7%
Mogale City	792	5.8%	2.5%
Rand West	715	5.3%	2.0%
GAUTENG	13 616	100%	100%

Table 1.3: Sample composition by sex and population group for Gauteng province in the unweighted and weighted QoL 6 (2020/21) data

	Sex	Black African	Coloured	Indian/Asian	White	Other	Total
Unweighted	Male	37.2%	1.2%	0.8%	7.2%	0.1%	46.6%
	Female	43.1%	1.5%	0.7%	8.1%	0.0%	53.4%
	Total	80.3%	2.8%	1.5%	15.3%	0.1%	100%
Weighted	Male	40.5%	1.5%	1.6%	6.3%	0.1%	49.9%
	Female	40.0%	1.7%	1.5%	6.8%	0.1%	50.1%
	Total	80.4%	3.2%	3.1%	13.1%	0.2%	100%

Weight efficiency is a metric that determines the efficacy of the weighting algorithm. The weighting efficiency for the benchmarked weights was calculated for each ward in Gauteng. If the data for many respondents needs to be weighted heavily up or down, the efficiency percentage is low. Hence, a higher percentage is desired. A weight efficiency of at least 75% is used in practice as an acceptable value. An average weight efficiency for the GCRO QoL 6 calibration weights is 89.59% with a standard deviation of 9.24%.

REFERENCES

Cochran, W.G. (1977). *Sampling teniques, 3rd ed.* Wiley, New York.

Deville, J.C., and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, vol.87, pp.376-382.

Deville, J.C., Särndal, C.E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, vol.88, pp. 1013-1020.

Hamann, C. and de Kadt, J. (2021). *Quality of Life Survey 6(2020/21): Sample design.* Gauteng City-Region Observatory (GCRO). Johannesburg. Available online at: https://gcro.ac.za/research/project/detail/quality-life-survey-vi-202021/

Lohr, S.L. (2010). *Sampling: Design and Analysis, 2nd ed.* Brooks/Cole, Boston

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

Neethling, A. (2004). Calibration and integrated weighting in sample surveys. *Doctor of Philosophy dissertation*, University of the Witwatersrand, South Africa.

Neethling, A., and Galpin, J.S. (2006). Weighting of household survey data: A comparison of various calibration, integrated and cosmetic estimators. *The South African Statistical Journal*, vol. 40, 123-150.