WEIGHTING REPORT

GCRO QUALITY OF LIFE SURVEY 7 (2023/24)

OCTOBER 2024

Author:
Ariane Neethling

GCRO | Gauteng City-Region Observatory

The GCRO comprises a partnership of:

GAUTENG PROVINCIAL GOVERNMENT REPUBLIC OF SOUTH AFRICA

SOUTH AFRICAN LOCAL GOVERNMENT ASSOCIATION SALGA *Inspiring service delivery*

WITS UNIVERSITY

UNIVERSITY OF JOHANNESBURG

WEIGHTING REPORT

GCRO QUALITY OF LIFE SURVEY 7 (2023/24)

OCTOBER 2024

DOI: https://doi.org/10.36634/EOSA3095

Written by: Ariane Neethling

Design: Breinstorm Architects

Cover typesetting: Lumina Datamatics

Cover photographs: Tshepiso Seleka

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).

WEIGHTING REPORT

GCRO QUALITY OF LIFE SURVEY 7 (2023/24)

Contents

PREFACE

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

The Quality of Life (QoL) Survey has become the flagship project of the GCRO. The QoL Survey is designed to provide a regular understanding of the quality of life, socio-economic circumstances, satisfaction with service delivery, psycho-social attitudes, values and other characteristics of residents in Gauteng. It serves as a tracking and diagnostic tool, affording a rich information resource for those people in policy-making, business, civil society and the public wanting to see where progress is being made, and where concerns remain.

The QoL Survey is a household-based survey with randomly selected adults (18+ years of age) as respondents. The GCRO has conducted seven QoL surveys since its inception in 2009:

- QoL I (2009) with 5 836 respondents in Gauteng and a total of 6 636 across the wider Gauteng City-Region (GCR).
- QoL II (2011) with 16 729 respondents in Gauteng.
- QoL III (2013/14) with 27 490 respondents in Gauteng.
- QoL IV (2015/16) with 30 002 respondents in Gauteng.
- QoL V (2017/18) with 24 889 respondents in Gauteng.
- QoL 6 (2020/21) with 13 616 respondents in Gauteng.
- QoL 7 (2023/24) with 13 795 respondents in Gauteng.

This publication is one of a series of technical reports about QoL 7 (2023/24). The reports include the Questionnaire, Fieldwork Report, Data Report, Sampling Report and the Weighting Report, as well as a generic guide to weighted analysis. These reports go hand in hand with the public dataset and should be consulted when analysing the QoL 7 (2023/24) data.

Additional information on the QoL Survey can be found on the GCRO website.

Photograph by Tshepiso Seleka

1. INTRODUCTION

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

This document provides technical details on the processes followed to calculate weights for the GCRO Quality of Life Survey 7 (2023/24). Weights were calculated using adult population and household count estimates provided by GeoTerraImage (GTI) and Census 2022 results.

Weights should be used during analysis of the QoL 7 (2023/24) data as detailed in Quality of Life Survey 6 (2020/21): Analyses under complex sampling (Neethling, 2021).

1. DESIGN WEIGHTS

In order to obtain a representative sample of the population, a stratified multistage probability sample was designed with Ward 2020 as the stratification variable. The sample was designed within each stratum (ward) in three stages with the Enumerator Areas (EAs) as the primary sampling units, dwelling units as the secondary sampling units, and an adult person as the tertiary sampling unit. 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.

2.1 First stage – EA weight

In the first stage primary sampling units, i.e. EAs, were selected. For the calculation of the EA weight, it was assumed that the EAs were selected with equal probability, irrespective of the number of households and population within the EA. Therefore, the weight of an EA is calculated as the inverse of the inclusion probability of an EA, by

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

where

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

N_{PSU} is the number of EAs in the population in the stratum

2.2 Second stage – Household weight

From each selected EA, a predetermined number of households were selected with equal probability. The household weight per PSU is given by

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

where

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

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

2.3 Third stage – Person weight

In the final stage, a person aged 18 years or older was randomly selected from the drawn household to be interviewed. The respondent weight is calculated by

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

where n^*_{18+} is the average number of persons aged 18 years and older in the selected households in the EA. The average is used, instead of the observed number of persons 18+ in a household, to obtain more smooth design weights with less variation.

3. CALIBRATION

Finally, the design weights of the respondents, as calculated in section 2, were adjusted to compensate for differential non-response (i.e. under/over-representation of certain subgroups 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, rim weighting, and poststratification 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) population estimates and Census 2022 counts.

The calibration technique was applied per local or metropolitan municipality in Gauteng with race-by-gender and ward as auxiliary variables. This means that the final set of weights sums to the 18 years and older population total estimate per ward as well as race-by-gender numbers per municipality. Due to too few Coloured and Indian/Asian respondents in the local municipalities of Sedibeng and West Rand, the Coloured and the Indian/Asian males and females sum to the totals of these two district municipalities, instead of the individual local municipalities. The following table summarises the race-by-gender groups used in the calibration.

District Mun.	Local Mun.	Black African		Coloured		Indian/Asian		White	
Sedibeng	Emfuleni	Male	Female	Male	Female	Male	Female	Male	Female
Sedibeng	Midvaal	Male	Female					Male	Female
Sedibeng	Lesedi	Male	Female					Male	Female
West Rand	Mogale City	Male	Female	Male	Female	Male	Female	Male	Female
West Rand	Merafong City	Male	Female					Male	Female
West Rand	Rand West City	Male	Female					Male	Female
Ekurhuleni	Ekurhuleni	Male	Female	Male	Female	Male	Female	Male	Female
City of Johannesburg	City of Johannesburg	Male	Female	Male	Female	Male	Female	Male	Female
City of Tshwane	City of Tshwane	Male	Female	Male	Female	Male	Female	Male	Female

Since only 20 of the 13 795 respondents indicated population group Other, the dominant population group of people of the EA in which they were interviewed, were assigned to these individuals for the calibration process.

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. The average weight efficiency for the calibration weights is 82.72% with a standard deviation of 11.78%.

4. QoL 7 (2023/24) WEIGHT VARIABLES

The weighting variable 'DOWNSCALE_MUN_PP_BENCHWGT' provides an individual level weight, scaled to the QoL 7 (2023/24) sample size of 13 795, 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 (this is called 'pp_benchwgt_mun' in the dataset). 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.

Ariane Neethling (Dr) (PhD, Pr.Sci.Nat)
Professional Statistical Consultant

References

Bankier, M.D. (1988). 'Power allocation: Determining sample sizes for subnational areas'. *The American Statistician, 42*(3), 174–177.

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

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

Lehtonen, R. and Pahkinen, E.J. (1996). *Practical methods for design and analysis of complex surveys*. John Wiley & Sons.

Lohr, S. (2022). *Sampling: Design and analysis* [3rd edition]. Tayler & Francis.

Neethling, A. (2004). 'Calibration and integrated weighting in sample surveys'. PhD 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, 40*, 123–150.

About the author

Dr Ariane Neethling | ariane_neethling@yahoo.com

Dr Ariane Neethling is a highly experienced statistician and academic with over 30 years of experience in the field of statistics. She has made significant contributions as a lecturer and senior lecturer at renowned universities, including the University of Pretoria, Stellenbosch University and the University of the Free State. Dr Neethling's expertise lies in Sampling and advanced sampling, as well as multivariate statistics, and big data analysis. She has served as a member of the SA Statistical Council since 2013 and is registered as a Professional Natural Scientist with the South African Council for Natural Science Professions. She is also a member of the International Association of Survey Statisticians (IASS), the SA Statistical Association and the South African Marketing Research Association (SAMRA).

GAUTENG CITY-REGION OBSERVATORY

6th Floor University Corner
11 Jorissen St (Cnr Jorissen and Jan Smuts)
Braamfontein
Johannesburg
Gauteng, South Africa

tel +27 11 717 7280
email info@gcro.ac.za
www.gcro.ac.za