DATA REPORT

GCRO QUALITY OF LIFE SURVEY 7 (2023/24)

OCTOBER 2024

Authors:
Sthembiso Pollen Mkhize, Pedzisai Ndagurwa
and Shamsunisaa Miles-Timotheus

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

DATA REPORT

GCRO QUALITY OF LIFE SURVEY 7 (2023/24)

OCTOBER 2024

DOI: https://doi.org/10.36634/EOSA3097

Written by: Sthembiso Pollen Mkhize, Pedzisai Ndagurwa and Shamsunisaa Miles-Timotheus

Design: Breinstorm Architects

Cover typesetting: Lumina Datamatics

Cover photographs: Tshepiso Seleka

Copyright 2024 © 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).

DATA REPORT

GCRO QUALITY OF LIFE SURVEY 7 (2023/24)

Contents

Tables

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.

1. INTRODUCTION

This report provides an overview of the GCRO Quality of Life (QoL) 7 Survey (2023/24) dataset and describes aspects of the data collection process that have a direct bearing on the structure and quality of the analytical dataset. It also documents variable coding and recoding, derived variables included in the dataset and implementation challenges that may be relevant to the interpretation of survey data. The report should be reviewed in conjunction with the survey dataset, the questionnaire and the accompanying technical documentation, namely the Fieldwork Report (de Fortier and Loots, 2024), Sampling Report (Naidoo et al., 2024) and the Weighting Report (Neethling, 2024).

Please note that unless otherwise specified, all figures provided in this report are based on unweighted data. It is also worth noting that the QoL 7 (2023/24) dataset production was implemented in the Statistical Package for the Social Sciences (SPSS) software programme, therefore the coding for missing values may differ from how other statistical programmes treat them. This was aligned with Stata's treatment of missing values in the conversion of the primary SPSS dataset to the public dataset in Stata format.

1.1 Overview of dataset

The QoL 7 (2023/24) dataset comprises responses from 13 795 adults, sampled from all 529 wards in Gauteng province. Further detail on sampling processes is available in the Sampling Report (Naidoo et al., 2024). Details on the composition and distribution of the final sample are available in the Fieldwork Report (de Fortier and Loots, 2024). GeoSpace International was appointed as the QoL 7 (2023/24) data collection service provider through an open tender process. Data was collected through in-person interviews, from August 2023 through to April 2024. Some high-level information on data collection is provided in section 2 of this report, and more detailed information on data collection is available in the Fieldwork Report (de Fortier and Loots, 2024).

For most purposes, the weighted analyses of this dataset are the most appropriate. Three weighting variables are included in this dataset - one for households and two for individual respondents/ persons. The household weight is named **'HH_WEIGHT'** and is not downscaled.

Of the individual weights, one is named **'pp_benchwgt_mun'** and the other is, **'DOWNSCALE_MUN_PP_BENCHWGT'**. Both of these are benchmarked to ward-level adult population size, and municipal population by race and gender, and then downscaled to the sample size. The **pp_benchwgt_mun** weight variable is designed to reflect respective population counts in the metropolitan and district municipalities to give the total adult population size of the Gauteng province as provided by the 2022 census by Statistics South Africa.

The **DOWNSCALE_MUN_PP_BENCHWGT** weight adjusts the sample distribution based on the 2022 census without upscaling frequencies to mirror population counts. The GCRO recommends using the **DOWNSCALE_MUN_PP_BENCHWGT** weight for persons analysis, but researchers can use **pp_benchwgt_mun** if they prefer. Further information on the calculation of these weights is available in the Weighting Report (Neethling, 2024).

2. DATA COLLECTION BACKGROUND

2.1 Questionnaire development, structure and administration

The survey questionnaire was designed by the GCRO with input from academic experts and stakeholders in provincial government and the three metropolitan municipalities in early 2023. In response to recommendations from the Quality of Life Survey 10-year review (Orkin, 2020), the questionnaire has been maintained at a length similar to that of QoL 6 (2020/21) and is shorter than iterations from QoL V (2017/18) and earlier. Nonetheless, the retention of core questionnaire content (including demographics, core services, governance, life satisfaction, health, crime and safety, migration, etc) facilitates repeated cross-sectional comparisons and analyses across QoL survey iterations. In addition, new content has been included in QoL 7 (2023/24) on social cohesion, inclusive economies, COVID-19 and vaccine hesitancy. Content in other areas has been enriched. These are economic activities, sexual orientation and gender identity, and different forms of violence.

The final questionnaire includes 211 questions,[1] divided into 15 modules, as follows:

1. Administrative information;
2. Basic services;
3. Migration;
4. Community attitudes;
5. Transport;
6. Economic;
7. Governance;
8. Social mobility;
9. Life satisfaction;
10. Crime and safety;
11. Social cohesion;
12. Health (including COVID-19 and vaccine hesitancy);
13. Environment/Sustainability;
14. Demographic details; and
15. Experiences of violence.

Full questionnaire content is available in a separate document that accompanies this report.

Following participant selection and the informed consent process, sections 1 through 14 were administered by a trained fieldworker in a face-to-face interview. Responses were captured on a tablet using a Computer-Assisted Personal Interviews (CAPI) system – Kobo Toolbox. Due to the sensitive nature of questions in section 15, participants were asked whether they were willing to complete this section. Those who were willing self-administered this section on the data collection tablet, although a small proportion did request assistance from the interviewer. Please note that the responses collected in section 15 are not included by default in the public dataset due to their

[1] Fourteen (14) questions, which had Yes/No responses and were administered as binary variables, brought the total number of questionnaire items on the CAPI software to 448.

sensitivity. Researchers wishing to use this data must request it separately via the DataFirst website, with a proposal outlining their intended use.

The QoL 7 (2023/24) questionnaire was administered in all South African languages depending on the preference of the respondents, but for purposes of this data report, we refer to the English version of the questionnaire. In the PDF questionnaire accompanying this report, the English language question text for each variable is shown in standard font in the 'English questions' column of the questionnaire, and response options are provided in the 'English responses' column. Additional information shown to the fieldworker is displayed in the 'Fieldworker and training notes' column. To ensure the quality of the survey data, we also included a 'Quality control/coding notes' column to support quality assurance measures both during and after data collection.

2.2 Piloting and translation

A process of 'behind-the-glass' piloting of the draft questionnaire, in English, was undertaken in June 2023, which allowed for the testing and refinement of content. Following this, the questionnaire was translated into Sepedi, Xitsonga, isiZulu, isiXhosa, Setswana, Sesotho, isiNdebele, siSwati, Tshivenda and Afrikaans by professional translators at Better Language Company. These translations were reviewed by GCRO researchers fluent in each language, and adjusted where necessary, both to ensure that the meaning of the questions was accurately reflected in each translation, and also to ensure that the language used in translations would be easily and accurately understood by respondents.

The GeoSpace International field team also reviewed the survey instrument, along with all existing translations. Fieldworkers were trained on the English language questionnaire, as well as the translations relevant to their individual language profiles. During training, and a small scale in-field pilot in August 2023, feedback from fieldworkers and pilot respondents informed further small adjustments to the wording of some questions and translations. Please refer to the Fieldwork Report (de Fortier and Loots, 2024) for further detail on training and in-field piloting.

All translations were included in the electronic data collection application – Kobo Collect (see sub-section 2.3 below). Fieldworkers were able to select the respondent's language of choice at the start of each interview and were also able to move between languages during the course of the interview if needed. The main language in which the interview was completed is included in the final dataset as variable 'interview_lang'.

2.3 Collection of spatial data

Multiple satellite-based Global Positioning System (GPS) coordinates were collected for each completed interview. These include the coordinates of the visiting point in the database from which visiting points were sampled, the GPS coordinates captured during the in-field sampling processes and the GPS coordinates captured during and after interview completion. The GPS coordinates captured during in-field sampling were generally accurate, as these were always captured outdoors at the interview location. In areas such as complexes, estates and hostels, the sample selection coordinates were sometimes affixed to the entrance location rather than the selected dwelling unit, making these slightly further away from the interview point. GPS coordinates captured during interviews were sometimes less accurate or missing altogether when interviews were conducted indoors due to challenges with satellite connection. In such cases, quality-control spatial checks by

both the GCRO and GeoSpace International were used to determine that the interviews were conducted at the correct locations.

The interview location data (GPS coordinates) collected during fieldwork are not included in the final dataset in order to protect anonymity and confidentiality. Coordinates were used for quality assurance and control processes and to generate the spatial paradata, as detailed in sub-section 3.1. Researchers and analysts wishing to conduct spatial analyses should make use of these spatial paradata variables.

2.4 Data collection system

GeoSpace International used two software systems for the implementation of QoL 7 (2023/24) – Hexagon (HxGN) Smart Census and Kobo Toolbox. For purposes of field management, GeoSpace International customised the HxGN Smart Census system to meet project needs. On the backend, this supported the allocation of pre-selected visiting points to fieldworkers and provided various monitoring and progress dashboards. When installed on dedicated data collection tablets, the HxGN application provided fieldworkers with navigation instructions to the visiting point, and functionally to complete the in-field component of sample selection. Satellite-based GPS coordinates were recorded at various times during the use of this system in the field, operating independently of network coverage and with generally good accuracy.

The survey questionnaire was digitised and managed using the Kobo Toolbox system. The Kobo Collect application was installed on the data collection tablets and used to administer the interviews. The Kobo Collect application displayed question text on screen for the interviewer to read out, along with the relevant response options, and in some instances additional notes or information for the fieldworker. All content was shown in the selected interview language. During survey completion, interview duration (dur_mins) and GPS coordinates (for quality control) were recorded. The GPS coordinates were collected in two different ways. One way was manual collection at the start and end of the interview by the fieldworker, and the other was done automatically by the system in the background while the fieldworker was administering the questionnaire.

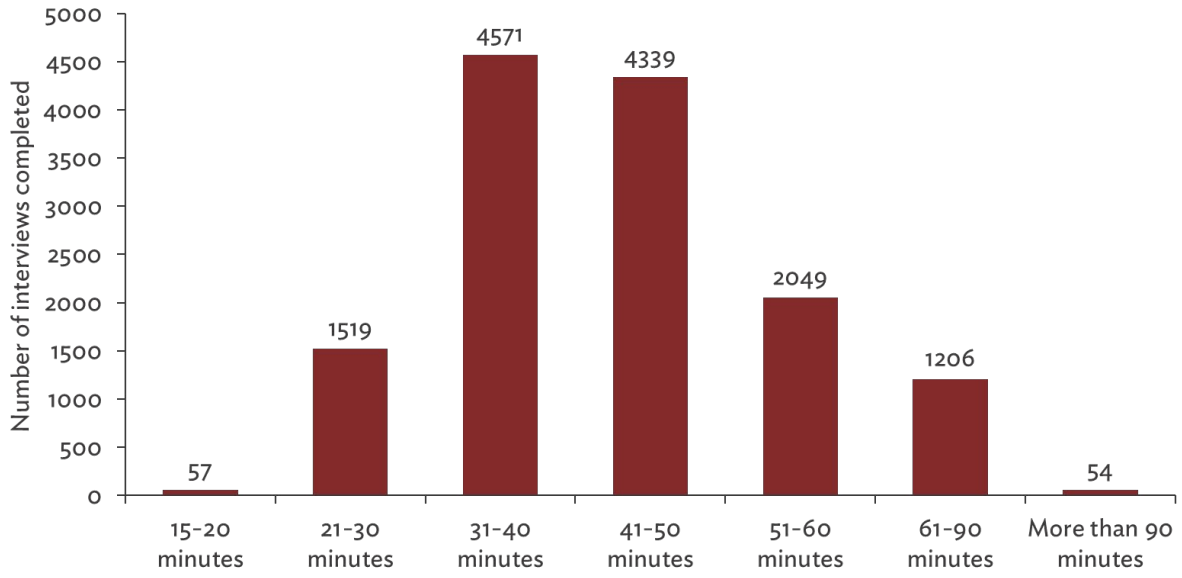
All aspects of both HxGN Smart Census and Kobo Collect functioned in the field regardless of network coverage. HxGN Smart Census and Kobo Toolbox were integrated such that each record's paradata (including GPS coordinates and sampling details) was linked with the relevant interview data. Further information on these systems is available in the Fieldwork Report (de Fortier and Loots, 2024).

2.5 Interview length

The average duration of the survey was 44 minutes for the fieldworker-administered segment and three minutes for the self-complete section. Figure 1 presents the distribution of interview length for the main part of the questionnaire. Informed by lessons from timed testing of the survey instrument in the pilot exercises, interviews with durations of less than 15 minutes were deemed unreliable and excluded from the survey dataset. Interviews with durations from 15 to 20 minutes (n=57; 0.4%) for QoL 7 (2023/24) were much lower than in the previous survey, QoL 6 (2020/21), but were still flagged by the quality assurance framework, and manually reviewed prior to a decision on whether they were reliable enough to include in the final dataset. Interview duration was also

monitored at the fieldworker and team levels and where these differed notably from average figures, this was investigated. Only 54 surveys (0.4%) lasted more than 90 minutes. These were largely caused by survey interruptions and were subsequently resumed, but also included some surveys where fieldworkers reported extremely talkative, slow or elderly respondents.

Figure 1: Duration in minutes of surveys in the final dataset



2.6 Question types

2.6.1 Fieldworker-administered questionnaire component

All questions in the fieldworker-administered component of the questionnaire required the selection of a response, unless they were automatically skipped (see sub-section 2.9 below). The main question types used in the fieldworker-administered section are described below.

Single-select: This is the default question type, offering a set of possible response options, of which only one can be selected. Examples include Q1.2c (rooms household occupies), Q1.5 (water meter) and Q1.7 (rainwater). After reading the question to the respondent, the fieldworker waited for the response and then captured this, repeating the question if necessary. For some questions, the response options were also read to the respondent. Where this was the case, this instruction was displayed on-screen for the fieldworker, and there is a note to this effect in the 'Fieldworker and training notes' column of the questionnaire. For some questions, a hard-copy show card was shared with respondents to assist them (see Q1.4 in conjunction with Annexure A of the Fieldwork Report). Where this was the case, the instruction to use the show card was displayed on-screen and is noted in the notes column of the questionnaire. Many of these questions made use of standardised Likert scales to reduce respondent burden. Details of these scales are included in sub-section 3.3 of this report.

Yes-No list: These questions provide a list of items below the question text, each of which was read to the respondent for a response of Yes or No, which was then captured by the interviewer. These questions are also treated as multi-select because they allow the respondent to choose 'Yes' for as many options as are applicable to them. Examples include Q1.12 (electricity supply), Q1.19 (extreme

events), Q5.14 (places walked to within 15 minutes from dwelling) and Q6.3 (household asset list). In Q1.12 (electricity supply), the final two items in the list ('Do not know' and 'No electricity'), were only displayed if the respondent had answered 'No' to all previous forms of electricity supply. Similarly, in Q10.5b (things done to look for work), the last item 'Don't know' was only displayed if the respondent had answered 'No' to all previous options.

Numerical and free-text: A small number of other questions required numerical input (including Q3.4 – year moved to Gauteng province; Q5.6 – travel duration; Q5.10 – money spent on food every month; Q5.11 – walking time to public transport; Q10.1b – number of people employed in a business; Q14.2 – age; and Q14.5, Q14.6 and Q14.7 – number of household residents). For a number of these variables, the dataset includes a categorical recode to facilitate analysis. Further information on these variables is available in section 4.1 of this report. Other response types included time (Q5.5 – departure time from dwelling to pursue frequent trip), free-text (Q10.10 – employment occupation/job title and 'Other – specify' questions as detailed in section 2.7), and a dropdown list (Q3.7 – neighbourhood of previous residence and Q10.1c – business location).

Additional comment spaces: Finally, throughout the questionnaire, there were a number of optional comment boxes allowing free-text inputs. These were used by fieldworkers if they wished to make a note of any information relevant to the responses captured, or if the respondent wished to highlight a particular issue for attention or reflect on their experience with the interview. All comments were reviewed during quality control (QC) and in finalising the analytical dataset, but they are not included in the final analytical datasets.

2.6.2 Self-complete questionnaire component

The self-complete section of the interview (section 15 in the questionnaire) made use of single response and Yes/No (binary) list question types. In this case, questions and response options were displayed on the tablet for only the respondent to read. Binary list questions included an instruction to select 'Yes' or 'No' to the applicable response options and a 'Prefer not to answer' option if the respondent did not feel comfortable answering a specific question. Where the respondent requested assistance, the fieldworker assisted by reading out the question text and response options.[2] Responses were required for all questions in this section, and skip patterns were not implemented.

2.7 Coding of free-text responses

Free-text responses to question Q10.10 (occupation), questions with option 'Not on the list' (which required respondents to describe their responses) such as Q3.7 and Q10.1c and those with the option 'Other (specify)' were reviewed, cleaned and coded as appropriate after data collection.

[2] Please note that this may have affected how respondents answered the violence questions if another person was present in the household. For instance, if the respondent was a victim and the perpetrator was in the same room, they might have been uncomfortable revealing sensitive information during the interview if the questions were read out loud.

2.7.1 Previous neighbourhood of residence in Gauteng (Q3.7) and business location (Q10.1c)

Responses to this question were manually checked, and spelling errors were corrected. These questions comprised a searchable dropdown list featuring area names of all neighbourhoods in Gauteng. Respondents were also given the option to select 'Not on the list' if the area name was not listed as a response option. This option was treated similarly to the 'Other (specify)' response, which enabled fieldworkers to input the correct area name as spelled out by the respondent.

2.7.2 Employment occupation (Q10.10)

Respondents who did some type of work, business or activity for pay (Q10.2) or reported that they had been appointed to a new job even though they had not started (Q10.3) were asked to describe their usual occupation using no more than two words. The responses were manually reviewed and coded in alignment with the major and sub-major occupation codes specified in the South African Standard Classification of Occupations (SASCO) (Statistics South Africa, 2012). Original responses are included in the dataset as variable '**q10_10_occupation_orig**'. Major categories are provided in the variable '**q10_10_occupation_maj**', and sub-major in the variable '**q10_10_occupation_submaj**'. In a small number of instances (n=11), indicated as -99, the responses provided by respondents did not fit within the predefined SASCO categories.

2.8 Use of 'Other' and 'Other (specify)' response options

Some single response and Yes-No list questions include an 'Other' response option for use when the predetermined standard options are not appropriate. Fieldworkers were trained on how to minimise the use of 'Other' fields by asking for more detail if necessary.

The 'Other (specify)' option was available in a subset of questions based on findings from the pilot exercise and, in some cases, was informed by lessons from previous survey iterations where questions had high proportions of 'Other' responses. When the 'Other (specify)' response option was selected, a free-text field was displayed, and the fieldworker captured further information as provided by the respondent. At the end of data collection, free-text responses in 'Other (specify)' fields were reviewed by dedicated QC team members at GeoSpace International. Where appropriate, these responses were recoded into one of the existing categories. Where the details did not merit recoding into an existing category, or where numbers were insufficient to introduce a suitable new category, the response remained coded as 'Other (specify)' in the analytical dataset. For the protection of anonymity and confidentiality, sensitive free-text responses to 'Other (specify)' are not shared with the dataset. Details of the response patterns and recoding are listed in sub-section 2.8.2 below.

2.8.1 Use of 'Other'

Table 1 lists all questions in which an 'Other' response option was available. It also provides the percentage and count of those who selected 'Other'. In most instances, less than 2% of respondents answering any question made use of 'Other'. However, there were four questions in which more than 2% responded with 'Other'.

Table 1: Availability and use of 'Other' response options across the questionnaire

Question	Percentage (count) 'Other'
Population group (A1)	0.1% (n=20)
Dwelling tenure (Q1.3)	1.0% (n=139)
Ownership type (Q1.3a)	*1.6% (n=139 of 8 646)
Rental type (Q1.3b)	*1.2% (n=41 of 3 415)
Main water source (Q1.4)	0.3% (n=36)
Main toilet type (Q1.10)	0.1% (n=8)
Refuse removal (Q1.11)	0.5% (n=65)
Energy for cooking (Q1.15)	0.4% (n=49)
Energy for lighting (Q1.16)	2.1% (n=290)
Type of dwelling moved from (Q3.8)	*0.8% (n=80 of 10 092)
Mode of transport used (Q5.7)	*0.4% (n=54 of 13 616)
Previous mode of transport used in the last 12 months (Q5.8)	*1.1% (n=156 of 13 616)
Debt (Q6.2.9)	*5.8% (n=804)
Things done to look for work or start a business (Q10.5b)	*4.9% (n=188 of 3 832)
Reasons for not seeking medical care (Q13.16)	*6.4% (n=13 of 203)
Reasons for not taking a COVID-19 vaccine (Q13.18)	*6.8% (n=337 of 4 929)
Reasons for taking a COVID-19 vaccine (Q13.19)	*0.6% (n=49 of 8 866)
Relationship status (Q14.4)	0.2% (n=28)
Home language (Q14.9)	3.8% (n=528)

**Note, percentage figures are the percentage of those answering the question, and not the percentage of the full sample.*

2.8.2 Use and recoding of 'Other (specify)'

Table 2 lists all questions which made use of an 'Other (specify)' response option. The table provides information on the percentage and counts of respondents whose responses were captured under 'Other (specify)'. No new response categories were required for the variables, as additional categories were added in the previous survey, QoL 6 (2020/21). In the processing of the dataset and through QC procedures, some responses initially captured under 'Other (specify)' were coded back to pre-existing categories.

The implementation of Q5.1 and Q5.3 as a consolidated variable (q5_1_3_frequent_trip_cons) was relatively easier compared to QoL 6 (2020/21).

A question on population group (Q15.2a) was included in the self-complete section to allow respondents to select their own race. This decision was made because the population group (a1_pop_group) in the main questionnaire is based on the race that the fieldworker perceives the respondents to belong to. The 'Other (specify)' responses in Q15.2a included 13 respondents (0.01%).

Table 2: Availability of 'Other (specify)' after recoding

Question	Percentage (count) of 'Other (specify)'	
	Original data	Final data (after recoding some of the 'other (specify)' responses
Dwelling type (A3)	0.1% (n=13)	N/A
Reason for municipal arrears (Q1.18)	*1.7% (n=30 of 1 730)	N/A
Previous Neighbourhood (Q3.7)	*0.7% (n=54 of 7 996)	N/A
Reason moved to neighbourhood (Q3.9)	*3.1% (n=109 of 10 092)	N/A
Biggest community problem (Q4.2)	0.3% (n=41)	N/A
Travel purpose (Q5.1)	0.5% (n=72)	0.1% (n=20)
Reason didn't vote (Q7.3)	0.5% (n=71)	0.4% (n=49)
Other clubs or organisations (Q12.1)	0.3% (n=41)	N/A
Reason for protests (Q12.4)	0.5% (n=74)	0.1% (n=19)
Petitioning (Q12.6)	0.2% (n=31)	*0.0% (n=1)
Reason for non-use of public healthcare services (Q13.2)	*1.1% (n=41 of 3755)	*0.7% (n=28 of 3755)
Household structure (14.13)	0.4% (n=52)	N/A
Race (Q15.2)	*0.2% (n=18 of 10 543)	*0.1% (n=11 of 10 543)

**Percentage figures are the percentage of those answering the question (applicable universe), and not the percentage of the full sample.*
N/A in this Table means that implementation of other specify responses were not required after recoding.

2.9 Skip patterns

A number of skip patterns were used in the questionnaire to avoid asking respondents questions that were not applicable to them. All skip patterns are documented in the survey questionnaire and are also listed here. Where a particular response option triggers a skip, this is indicated in brackets

after the response option. Where a particular question is only asked when particular conditions are met (or inversely, skipped under certain conditions), this is specified in the 'Notes' column. When a question was not asked of a respondent, the dataset makes use of a '–1'[3] to indicate that the response is missing due to a skip pattern. Further information on coding of missing data is available in section 3.2 of this report.

2.9.1 Main questionnaire

Skip patterns were used in the following questions in the fieldworker-administered questionnaire.

Number of households sharing a room (Q1.2c & Q1.2d): Q1.2c asked respondents how many rooms their household occupies. Respondents who indicated one room were then asked how many other households share this room, as a measure of crowding. When the household occupied more than one room, Q1.2d was not asked.

Household tenure (Q1.3, Q1.3a & Q1.3b): Respondents were asked about dwelling tenure in Q1.3. If respondents indicated that the dwelling was owned by the household, they were asked for further ownership details in Q1.3a. If respondents indicated that the dwelling was rented, they were asked for further details on the rental arrangement in Q1.3b. When any other option was selected in Q1.3, these additional questions were both skipped.

Main water source (Q1.4, Q1.5, Q1.6 & Q1.7): All respondents were asked Q1.4, about the household's main water source. When respondents indicated that they received piped water, either into the dwelling or into the yard, they were additionally asked about the type of water meter the household has (Q1.5). Those who selected any other response option were not asked Q1.4. Additionally, respondents who answered 'Well or borehole' to Q1.4 were not asked Q1.6 (whether the household also gets water from a well or borehole), and those who selected 'Rainwater tank (e.g. a Jojo tank)' were not asked Q1.7 (whether the household also gets water from a rainwater tank).

Electricity supply (Q1.12): This question asked respondents to identify all forms of electricity supply that their household made use of, and was implemented as a Yes-No list (see sub-section 2.6.1 above for more information). Item 2 ('Electricity with post-paid meter') was only displayed and read out if the respondent selected 'No' to item 1 ('Electricity with pre-paid meter'). Additionally, the final two options ('Do not know' and 'No electricity') were only displayed and read out to the respondent if they had answered 'No' to all of the seven preceding items. Respondents who answered 'No' to all alternative energy sources (Solar water heater; Solar PV panels; Inverter/batteries; Petrol or diesel generators; Wind power (e.g. windmills)) were not asked Q1.12a.

Investing in alternative energy sources (Q1.12b and Q1.12c): All respondents were asked Q1.12b, which presented a Yes-No list, to determine whether their household was considering investing or further investing in alternative energy sources. Respondents who answered 'No' to all options were not asked Q1.12c, as it pertains to the main reason for wanting to invest in alternative energy sources.

[3] As noted at the end of section 1 (Introduction), this is for the SPSS dataset. The missing values were transformed to the applicable coding in the Stata version of the dataset.

Electricity supply interruptions (Q1.14): Respondents were only asked Q1.14, about how frequently they experienced electricity interruptions in the past 12 months, if they had indicated that they have access to electricity that is not off-grid, that is, 'Electricity with pre-paid meter', 'Electricity with post-paid meter', 'Connection from neighbour's house' or 'Connection from elsewhere (including direct to overhead cables)' in Q1.12.

Reason for arrears in water or electricity accounts (Q1.17 and Q1.18): Question 1.17 asked respondents whether they had unpaid accounts for services like water and electricity. Respondents who answered 'Yes' to Q1.17 were asked Q1.18 (main reason for arrears), while all others skipped to Q1.19.

Timing of move to Gauteng (Q3.1, Q3.3 & Q3.4): In Q.3.3, all respondents were asked to select the province or the option 'country outside South Africa' where they had lived most recently before coming to Gauteng. Those who indicated in Q3.3 that they have 'Always lived in Gauteng' were not asked Q3.4. If respondents who indicated that they were born outside of Gauteng selected 'Always lived in Gauteng' in Q3.3., a logic check was triggered (see section 2.10.1).

Moving into current neighbourhood (Q3.5): All respondents, regardless of place of birth, were asked how long they had lived in their current neighbourhood. Those who indicated that they had always lived in that neighbourhood (indicated as '9999' in the dataset) skipped through to Q3.9a, while all other respondents continued to answer additional questions about the move to their current neighbourhood. Seven responses were marked as –3 'fieldworker error' after a logic check during QC revealed discrepancies between responses recorded for Q3.3 and Q3.4. These two frequencies should have aligned, as a response of –1 indicated that the respondent has always lived in Gauteng. Therefore, the seven respondents who selected this option should not have been able to answer Q3.4.

Why respondents came to the current neighbourhood (Q3.6, Q3.7 & Q3.8): These questions were administered to respondents who previously lived in another neighbourhood, based on their response to Q3.5 (year moved into the current neighbourhood). In Q3.6, respondents were asked whether they had moved to their current neighbourhood from somewhere else in Gauteng, another province or another country. Only those respondents who had moved from somewhere else in Gauteng were asked Q3.7, which asked for the details of the neighbourhood that they had moved from. Those who had moved to their current neighbourhood directly from another country or province skipped through to Q3.8, which asked about the dwelling type they had moved from.

Purpose of most frequent trip (Q5.1, Q5.2 & Q5.3): Question 5.1 asked respondents the purpose of their most frequent trip. If a respondent answered that they didn't make any trips, this was recorded, but they were then asked Q5.2, which asked them if they were sure that they never went anywhere. A small number of respondents (n=19) answered 'No' (meaning they do indeed go somewhere) to this question. These individuals who do go somewhere were then asked Q5.3, which offered the same response options as Q5.1. We have provided a variable which includes information on all trip purposes, whether captured in Q5.1 or Q5.3 – more information is available in sub-section 4.2.5. Individuals who confirmed in Q5.2 that they really never went anywhere skipped through to Q5.14.

Details of most frequent trip (Q5.5, Q5.6, Q5.7 & Q5.8): Question 5.5 through to Q5.8 were only answered by respondents who reported a trip in Q5.1 or Q5.3.

General transportation information (Q5.10, Q5.11, Q5.12 & Q5.14): Of these questions, only Q5.14 (places walked to within 15 minutes from dwelling) was asked of respondents who reported that they never went anywhere in Q5.2.

Debt (Q6.2 & Q6.2a): All respondents were asked in Q6.2 about whether they currently had different types of debt (Yes-No list). Those who responded 'Yes' to any of the options were asked Q6.2a, which asked whether they had missed a debt repayment in the past three months. Those who answered 'No' to all of the options in Q6.2 skipped through to Q6.3.

Child hunger (Q6.5 & Q6.5a): All respondents were asked Q6.5, which asks whether there has been a time in the past 12 months where there was not enough money to feed children in the household. Response options were 'Yes', 'No' and 'There are no children in the household'. Those who said that there were no children in the household were not asked Q6.5a (whether children benefited from school feeding schemes) but skipped through to Q6.8. Those who answered 'Yes' or 'No' to Q6.5 were asked Q6.5a before continuing to Q6.8.

Voter registration and electoral participation (Q7.1, Q7.2 & Q7.3): Question 7.1 asked respondents whether they were a registered voter. Those who said 'No' skipped through to Q7.4. Those who said 'Yes' in Q7.1 were asked Q7.2 (whether they plan to vote in the 2024 national/provincial elections). Response options were 'Yes', 'No' and 'Not sure'. Those who answered 'Yes' or 'Not sure' were not asked Q7.3 (the main reason that they do not plan to vote).

Business ownership (Q10.1 - Q10.1e): In sub-section Q10.1, respondents were asked about business ownership. Q10.1 asked whether they currently owned a business, regardless of whether they started, purchased, or inherited it, even if they are no longer actively involved in its operations. Respondents who answered 'No' proceeded to sub-section Q10.2.

Currently working (Q10.2, Q10.3): All respondents were asked whether they had done any type of work in the past seven days (Q10.2). Those who said 'Yes' were categorised as working and skipped through to Q10.6. Those who said 'No' were then asked Q10.3 (whether they had been appointed to a new job but had not yet started). Those who said 'Yes' to Q10.3 were additionally categorised as working and skipped through to Q10.6, while those who said 'No' to Q10.3 were categorised as not working and were asked Q10.4.

Unemployment (Q10.4, Q10.5, Q10.5a & Q10.5b): All non-working individuals (see preceding paragraph) were asked if they were unemployed and looking for work (Q10.4). Those who answered 'No' in Q10.4 were then asked why they were not looking for work (Q10.5), and after answering this question, they then skipped through to Q11.1. Those who responded 'Yes' skipped through to Q10.5a, which asked how long it had been since they last worked and proceeded to be asked Q10.5b, which asked what they had done to look for work or start a business in the past four weeks. After answering these questions, all non-working respondents were directed to skip to Q11.1.

Employment details (Q10.2, Q10.3, Q10.6, Q10.8, Q10.9 & Q10.10): Only the individuals categorised as working (see above) in Q10.2 or Q10.3 were asked questions Q10.6 through to Q10.10, which collected further information about the nature of the respondent's work.

Healthcare services used (Q13.1, Q13.2, Q13.3 & Q13.4): All respondents were asked in Q13.1 where they usually went for healthcare. Respondents who answered 'Not applicable, don't usually need healthcare' were not asked any further questions about health services used and skipped through to Q13.6. Individuals who indicated in Q13.1 that they only used public healthcare facilities, or a combination of public and private healthcare facilities were not asked Q13.2 but skipped through to answer Q13.3 and Q13.4. Those who answered in Q13.1 that they used only private healthcare facilities, traditional healers or spiritual healers were asked in Q13.2 for the main reason that they didn't make use of public healthcare facilities. They were then also asked Q13.4 onwards.

COVID-19 and vaccination (Q13.15–Q13.20): In Q13.15, respondents were asked about their experience the last time they dealt with COVID-like symptoms, with various response options made available. If respondents did not select the response 'I thought I had COVID-19 but did not seek medical care', they skipped Q13.16 (reason for not seeking medical care) and proceeded to answer Q13.17. In Q13.17, respondents were asked whether they had received a COVID-19 vaccine. Those who answered 'No, I did not' were asked Q13.18, while others skipped to Q13.19 (reason for getting the vaccine or intending to get one). In Q13.20, respondents were asked whether it was easy or difficult to get the most recent vaccine. However, those who answered 'No' or 'No, but I intend to' to Q13.17 were not asked this question.

Number of people in household (Q14.5, Q14.6 & Q14.7): In Q14.5, respondents were asked for the number of people, including babies and children, living in the household. If respondents answered '1' in Q14.5, they skipped Q14.6, which asks for the number of residents under the age of 18 because this meant that the household did not have any member below 18. If the respondent indicated more than one resident, they were asked for the number of residents under 18 in Q14.6, and the number of residents aged 65 or more in Q14.7. Validation was applied to responses, as detailed in sub-section 2.10 below.

2.9.2 Self-complete module

Due to the sensitive nature of the questions in section 15 (the self-complete module), a second consent process was conducted to ensure respondents agreed to participate prior to being offered the tablet to complete this section confidentially. If a respondent did not consent to participate in this section, the interview was ended at this point and sent for quality control, and all questions in the self-complete section were set as missing due to skip (–1). The variable 'sc_consent' indicates whether or not the respondent consented to participate in this section, with '1 = Yes' and '0 = No'. For respondents who did complete the section, a 'Prefer not to answer' response option was available for every question. In most instances, selection of this option would simply move the respondent to the next question. However, in one instance, a pop-up note (serving as a logic check) appeared when the respondent was asked about their party choice if they were asked to vote in elections (Q15.10). This occurred because they selected 'Would not vote' in Q15.10 but indicated 'Yes' to planning to vote in Q7.2.

Please note that the variables in the self-complete module are not included in the publicly available survey dataset and must be requested separately from DataFirst.

2.10 Logic checks, validations and value constraints

A small number of logic checks and validations were built into the questionnaire, which were used during data collection to provide fieldworkers with an opportunity to confirm these responses. A number of fields requiring a numerical response also included built-in value constraints, to reduce errors in entering numbers. These logic checks, validations and value constraints are documented below. As previously indicated, all questions in the main questionnaire required a response, unless skipped by a skip pattern. It was not possible for the fieldworker to continue to the next question unless all previously asked questions had responses recorded.

As part of the Quality Assurance (QA) process, a further series of automated data checks were run on all incoming survey responses to identify anomalous responses. Where these data checks identified issues, surveys received additional manual scrutiny by the QA team at GeoSpace International. These automated checks are detailed in sub-section 2.11.3 below.

2.10.1 Logic checks and validations built into the questionnaire

The following checks were built into the survey questionnaire and used during data collection.

Household ownership and renting (Q1.3 & Q1.3a): If a respondent selected 'Squatting or living rent-free in an informal dwelling they built, or in a vacant building or on vacant land' in Q1.3, but also indicated in Q1.3a that they lived in a 'Free RDP house', it was not possible to continue to the next question until this was resolved.

Water access (Q1.4): If the fieldworker selected 'House, brick or concrete structure on a separate stand', 'Flat or apartment in a block of flats', 'Cluster house in a complex', 'Townhouse (semi-detached house in a complex)', 'Semi-detached house not in a complex', 'House, flat or room separate from main dwelling in backyard', 'Room or flat which is part of main dwelling or property', 'Unit in a retirement home or barracks' or 'Hostel', it was considered very uncommon for the respondent to choose 'Flowing river or stream' or 'Dam, pool or standing water' as their main water source. A pop-up note appeared to indicate that the questionnaire could not proceed until the response from the respondent was confirmed.

Water cleanliness (Q1.8): If the respondent had previously indicated in Q1.4 that their household received piped water into the dwelling, and then indicated in Q1.8 that the water they received was hardly ever or never clean, this was considered very uncommon. The questionnaire could not proceed until this inconsistency was acknowledged by allowing the fieldworker to click 'OK' on the pop-up note that appeared after the respondent had confirmed this as true.

Toilet type (Q1.10): If the respondent had previously indicated that their dwelling type (A3) is 'Flat or apartment in a block of flats', 'Cluster house in a complex', 'Townhouse (semi-detached house in a complex)' or 'Semi-detached house not in a complex', and then indicated in Q1.10 their household has access to a 'Chemical toilet', 'Pit latrine with ventilation pipe', 'Pit latrine without ventilation pipe', 'Bucket toilet', 'Communal toilet', 'Neighbour's toilet', 'No access to toilet' or 'Other', this was considered very uncommon. The questionnaire could not proceed until this inconsistency was acknowledged by allowing the fieldworker to click 'OK' on the pop-up note that appeared after the respondent had confirmed this as true.

Electricity supply (Q1.12): In this Yes-No list, it was not possible for respondents to respond 'Yes' to both items: 1 ('Electricity with pre-paid meter') and 2 ('Electricity with post-paid meter'). This was achieved by not displaying item 2 unless 'No' was recorded for item 1. Where item 2 was not displayed, this variable was coded as 'missing due to skip' ('–1').

Similarly, it was not possible for respondents to respond 'Yes' to item 8 (Do not know) or item 9 ('No electricity') along with any of items 1–7. This was achieved by not displaying items 8 or 9 unless a 'No' had been recorded for all previous items. If items 8 and 9 were not displayed, they were coded as 'missing due to skip' ('–1').

Electricity interruptions (Q1.14): In Q1.14, respondents were asked how often they had experienced electricity interruptions (excluding planned outages and load shedding) in the past 12 months. Respondents who reported no access to electricity and alternative energy sources in Q1.12 were not asked this question. At the start of fieldwork, Q1.14 was only posed to those who answered 'Yes' to either 'pre-paid' or 'post-paid' electricity in Q1.12. However, starting from 18 October 2024, this question was extended and made available to all respondents who selected 'Yes' to the initial two categories, as well as those who answered 'Yes' to accessing electricity from a neighbour, elsewhere or 'don't know' in Q1.12. This change provides researchers with greater flexibility in analysing this data. Analysts using Q1.14 should be aware of these discrepancies and ensure they make informed decisions and conduct thorough scrutiny before using this variable.

Energy source used for cooking (Q1.15): In Q1.13c, respondents were asked whether they had to cook using alternative energy sources such as gas, paraffin or coal instead of electricity due to loadshedding. Respondents who selected 'Not applicable, this household hasn't used electricity for cooking' in Q1.13c and those who previously indicated that there was 'No electricity' in the household in Q1.12 were unable to choose 'Electricity' in Q1.15. If they did attempt to select 'Electricity', a pop-up note appeared to remind the respondent about their response in Q1.13c. It was not possible to continue until this issue was resolved.

Energy source used for lighting (Q1.16): Similar to the above, in Q1.16, respondents were asked about the most used energy source for lighting in their household. If a respondent selected 'Solar/renewable' as the most used energy source, a pop-up note appeared if they had not indicated access to any solar energy source in Q1.12, or if they had indicated 'No electricity' in Q1.12 but selected 'Electricity' in Q1.16. It was not possible to continue until this issue was resolved.

Place of birth and duration of stay in current neighbourhood (Q3.1 & Q3.3): If a respondent indicated in Q3.1 that they were born outside of Gauteng province, and then in Q3.3 indicated that they had 'Always lived in Gauteng', an alert was displayed on the screen to ask the fieldworker to review these responses. It was not possible to continue until this was resolved.

Year moved to Gauteng and in the current neighbourhood (Q3.4 & Q3.5): In Q3.4, respondents were asked about the year they had moved into Gauteng, while Q3.5 asked about their move into their current neighbourhood. If the year captured in Q3.5 preceded the year captured in Q3.4, this was flagged as logically inconsistent. It was not possible to continue until this was resolved.

Most frequent trip (Q5.1, Q5.2 & Q5.3): In Q5.1, if a respondent selected the response option 'I don't make any trips', this was verified by asking the respondent to confirm that they never left to go

anywhere in Q5.2. If the respondent confirmed that this was correct, they skipped all related questions. If a respondent indicated that they did in fact go somewhere, these details were collected in Q5.3 and the respondent completed all related questions.

Reason for not planning to vote (Q7.3): In Q3.1, if a respondent specified that they had been born in a province in South Africa, but then noted in Q7.3 that their reason for not planning to vote was because they were not a South African citizen (both of these responses cannot be true), the respondent could not proceed to the next question unless they confirmed the correct response to either Q3.1 or Q7.3.

Satisfaction with different levels of government (Q7.4, Q7.5 and Q7.6): Fieldworkers were advised to check if there was a typing/finger error if the respondent had indicated that they were very dissatisfied with government initiatives to grow the economy and create jobs (Q2.11) or very dissatisfied with the government's response to COVID-19 (Q2.12), but then selected very satisfied with the performance of either level of government. If it was not a typing/finger error, the respondent was asked to confirm whether their responses were true and correct.

Satisfaction with marriage or relationship (Q9.4): If the respondent was not married or in a relationship, their answer to Q9.4 would typically be 'Not applicable'. The pop-up note appeared when the respondent selected the other given options and was used to confirm whether the respondent was indeed married or in a relationship.

Main reason for not using public health facilities and being covered by medical aid (Q13.2 & Q13.5): In Q13.2, if the respondent indicated that they did not use public health facilities because they had private medical aid cover but did not select 'Yes' to being covered by medical aid in Q13.5, this was considered incorrect. It was not possible to proceed until this was corrected.

Health status (Q13.6, Q13.7 & Q13.8): It was considered very uncommon if the respondent indicated that their state of health in the past four weeks had been excellent (Q13.6), but also noted that their health status 'Always' prevents them from doing daily work (Q13.7) or partaking in social activities (Q13.8). It was impossible to proceed to the next questions without rectifying this inconsistency.

Highest education level and respondent's age (Q14.1 & Q14.2): If the respondent indicated that they had completed postgraduate tertiary education in Q14.1, but their age was less than 21 (Q14.2), this was considered very uncommon. The fieldworker had to confirm if this was indeed true and correct before proceeding to the next question.

Respondent's age and year moved to Gauteng (Q14.2 and Q3.4): If it was indicated that the year that the respondent had moved to Gauteng (Q3.4) was before the year they were born (Q14.2), a pop-up note appeared. It was not possible to continue until this was corrected.

Number of household residents (Q14.5, Q14.6 & Q14.7): After answering Q14.6 (number of residents under 18) and Q14.7 (number of residents aged 60 or above), respondents were asked to confirm their response prior to continuing to the next question. If the response to either Q14.6 or Q14.7 exceeded the value provided in Q14.5 (total number of household residents), an alert was displayed on the screen, and it was not possible to proceed until this was corrected. Additionally, if

the sum of Q14.6 and Q14.7 exceeded the answer provided in Q14.5, an alert was displayed on the screen, and it was not possible to continue until this was corrected.

Household head (Q14.8): If the respondent reported that only one person lived in the household in Q14.5, they could only select 'Respondent' in Q14.8. If the respondent selected an option other than 'Respondent', an error message popped up reminding them that they had reported that they were the only person living in the household in Q14.5. The respondent was given the opportunity to either correct the response for Q14.5 if they had made a mistake or to select 'Respondent' in Q14.8 as the correct response.

Statement best describing the household (Q14.13): Similar to the above, Q14.13 asked respondents the best option that described their household. If the respondent reported that only one person lived in the household in Q14.5, they could only select 'Single-person household/only me' in Q14.13. If the respondent selected an option other than 'Single-person household/only me', an error message popped up reminding them that they had reported that they are the only person living in the household in Q14.5. The respondent was given the opportunity to either correct the response for Q14.5 or select 'Single-person household/only me' in Q14.13 as the correct response.

Household income (Q15.3): A validation prompt appeared if the respondent reported 'No income' in Q15.13 to ensure that there was truly no income coming into the household. Additional validations were implemented in Q15.3 if the respondent reported some form of income in Q5.10, Q6.8 or Q14.12.

Party choice (15.10): A logic check was implemented in Q15.10 to address cases where a respondent selected 'Would not vote' but answered 'Yes' to planning to vote in Q7.2. The respondent could not submit the self-completed questionnaire unless this inconsistency had been resolved in either Q7.2 or Q15.10.

2.10.1 Value constraints

Value constraints were applied to a number of questions requiring numerical responses to limit the chance of an invalid number being erroneously captured. These constraints are detailed in the questionnaire, and below.

- **Year moved to Gauteng (Q3.4):** For interviews conducted in 2023, the value could not be more than 2023, and for those conducted in 2024, it could not be more than 2024.
- **Duration of most frequent trip in minutes (Q5.6):** The value entered could not be less than 0.
- **Transport expenditure (Q5.10):** The value entered could not be more than five digits.
- **Public transport proximity (Q5.11):** The value entered could not be less than 0.
- **Food expenditure (Q6.8):** The value entered could not be more than five digits.
- **Year of business ownership (Q10.1a):** For interviews conducted in 2023, the value could not be more than 2023, and for those conducted in 2024, it could not be more than 2024.
- **Respondent age (Q14.2):** The value entered could not be less than 18 and could not be more than 115. All ages over 80 were verified during QA. Please note that only a recode of the age variable is included in the public dataset in order to protect the anonymity of participants.

- **Number of people living in household (Q14.5):** The value entered could only be between 1 to 30.
- **Number of people under 18 living in the household (Q14.6):** The value could not be more than the response recorded in Q14.5.
- **Number of people aged 65 or more living in the household (Q14.7):** The value could not be more than the responses recorded in Q14.5.

2.11 Quality assurance processes

Quality assurance (QA) processes were jointly developed by GeoSpace International and the GCRO. The implementation of these processes was done by GeoSpace International. Key sampling, spatial and dataset QA processes are documented here, and further information on QA carried out by GeoSpace International is also available in the Fieldwork Report (de Fortier and Loots, 2024).

The general QA workflow involved a series of automated checks run by GeoSpace International on all incoming interviews, as detailed below. Interviews which were flagged by these automated checks received a manual review, which might trigger the confirmation of responses with the fieldworker or respondent, a return to field, correction of data errors or the rejection of the interview altogether. On a weekly basis, GeoSpace International also examined QA results at the level of the individual fieldworker and occasionally at the team level. The GCRO received completed interviews from GeoSpace International, with a record of QA results and progress for each questionnaire.

The GCRO independently ran a series of rigorous quality control (QC) checks complementing the QA processes to ensure that the GeoSpace International processes were working effectively. In addition, the performance of skip patterns and logic checks was reviewed on a weekly basis. Periodic manual checks were also used to look for any anomalies in the data, and a proportion of fieldworker and respondent comments were manually scanned. On detection of anomalies, the GCRO notified GeoSpace International in a shared Google sheet. GeoSpace International solved identified QC issues on a rolling basis, mainly using callbacks to respondents to verify and/or collect the correct responses.

2.11.1 Sampling and implementation checks

Linkage of HxGN Smart Census and Kobo Toolbox data: GeoSpace International ensured that each record on HxGN Smart Census was linked to exactly one completed questionnaire on Kobo Toolbox, and vice versa. Where this was not the case, the data was examined to enable manual matching.

Duplication of interview: GeoSpace International ensured that no records were duplicated through the examination of spatial, sampling and other paradata.

Linkage of main questionnaire and self-complete content: GeoSpace International ensured that there was exactly one valid self-complete record for each valid main questionnaire, and that these were linked correctly. (Note that even when the respondent chose not to complete the self-complete module, the fieldworker was obliged to submit a record of this decision).

Use of substitution points: Where interviews were conducted at substitution points, GeoSpace International ensured that the original pre-selected visiting points had been visited, with appropriate revisits as applicable, and that it had not been possible to conduct an interview at this point. GeoSpace International additionally ensured that the visits to the original visiting point preceded engagements with substitution points.

Number of visits per EA: GeoSpace International continuously monitored the number of interviews per enumerated area (EA) to ensure that the sample within each ward was appropriately distributed. This was particularly important in areas where there was extensive use of substitution.

Interview location: Throughout the data collection period, GeoSpace International validated the location of all interviews by ensuring interview coordinates were aligned with sample points. If a fieldworker completed an interview more than 50 meters from the original or substitute visiting point, a valid reason had to be provided. Where this was not the case, the interview was investigated. In many instances, discrepancies could be explained by the distance between the point at which an interview could be conducted and the original sample point. Where the difference could not be explained, the interview was not QA-accepted. Interview location was also validated by calculating the distance between coordinates collected throughout the interview. If there were notable distances between coordinates, this required investigation. The GCRO independently replicated interview location validation at various points during the course of data collection.

Appropriate respondent selection: The name, age and sex (and gender identity) of the sampled respondent in the household register was compared to information collected during the interview itself. Discrepancies were investigated. Interviews conducted with individuals other than the sampled adult were not accepted.

Interview language: Interview language was compared to reported home language. Where this differed, the fieldworker was asked to provide an explanation in a comments field. Explanations were manually reviewed.

Interview duration: Interviews with a main questionnaire duration of less than 20 minutes were all examined manually. Callbacks were made to double-check whether the interview took place and to determine its duration based on the respondent's input. Interviews with a main questionnaire duration of less than 15 minutes were QA-rejected. Self-completion durations were also further investigated if the respondent took less than one minute to complete the questions.

2.11.2 Questionnaire content checks

Number of households sharing a room (Q1.2c and Q1.2d): Where the respondent reported in Q1.2c that the household lived in a single room, and further indicated that the household shared this room with other households in Q1.2d, this information was verified, particularly if the number of other households was identical to the number of household members.

Number of adult residents (Q14.5): The number of adults listed in the household listing was validated against the number of adults indicated by responses to Q14.5 (total number of residents) and Q14.6 (total number of children).

Fieldworker comments: All interviews which included fieldworker comments were flagged, and the comments were manually reviewed by the GeoSpace International QA team. If necessary, further information was obtained from the respondent or fieldworker.

2.11.3 Automated fieldworker- and team-level checks

A number of checks were also examined at the fieldworker and team levels, as follows.

Duration: The number of interviews shorter than 20 minutes, average questionnaire duration and the distribution of durations were examined at the fieldworker level. Where fieldworkers had a high proportion of very short questionnaires or unusual average duration or distribution of durations, this was investigated further and remedied as appropriate.

Household listing (adult_count): The average number of adults listed for each household was examined to ensure that fieldworkers were listing adult residents appropriately. Where the average number differed notably from the mean across all fieldworkers, this was investigated further, with the provision of further training if needed. During QC, we found discrepancies between the number of adults reported in adult_count and those reported in Q14.5 and Q14.7. These discrepancies arose because some respondents were not willing to share the exact number of adults in their households during the roster. During callbacks, it was found that some respondents may have initially listed only one adult (themselves) during the roster due to the lack of trust in the survey. However, once these respondents began the questionnaire and realised it was legitimate, they disclosed the correct number of adults in the household. For analysis requiring the exact number of adults, researchers/analysts are advised to use Q14.5 in conjunction with Q14.6 by subtracting the total number of children reported in Q14.6 from the count provided in Q14.5.

Respondent sex (A2): The balance of male and female interviews conducted by each interviewer was monitored, and where this varied notably from an even split, this was investigated. In some instances, there were legitimate explanations for this, such as area demographics or fieldworker skill sets.

No problem (Q4.2): This was done to check how frequent fieldworkers and teams would select 'No problem'. Some fieldworkers may avoid this question to avoid waiting for the respondent to think about the main problem in the community.

Going nowhere (Q5.2): Quick questionnaires with a travel status of 'going nowhere' underwent additional checks via callbacks to ensure that fieldworkers and teams did not select 'Yes' in Q5.2 with the intention of skipping the entire transport section.

Business and work (Q10.1 & Q10.2): Questionnaires that had 'No' for Q10.1 or Q10.2 underwent additional checks via callbacks to ensure that fieldworkers and teams did not select these responses with the intention of completing the questionnaire quickly.

Health status and doing daily work (Q13.6 & Q13.7): This check was conducted to determine how often fieldworkers and teams selected 'Excellent' health status in Q13.6 and 'Always' for the respondent's health, preventing them from doing daily work in Q13.7.

Health status and engaging in social activities (Q13.6 & Q13.8): This check was conducted to determine how often fieldworkers and teams selected 'Excellent' health status in Q13.6 and 'Always' for the respondent's health, preventing them from participating in social activities in Q13.8.

Refusal rate for the self-complete module: The proportion of respondents refusing to complete the self-complete module was monitored. Where this differed substantially from the mean, this was examined. Further training was provided to fieldworkers who had very high refusal rates.

Refusal rate for provision of contact information: The proportion of respondents refusing to provide contact information was monitored. Where this differed substantially from the mean, this was examined and training offered if appropriate.

3. PARADATA, STANDARDISED CODES AND RESPONSE OPTIONS

This section provides an overview of the paradata included with the QoL 7 (2023/24) dataset, and documents the standardised codes used for missing data and frequently used response options in the dataset.

3.1 Spatial and other paradata

Paradata refers to variables in a survey dataset which describe the data collection process. For reasons of confidentiality, not all paradata related to QoL 7 (2023/24) can be made publicly available. Where possible, paradata variables are included in their raw form, but in other instances, we are only able to provide derived paradata variables.

3.1.1 Determination of survey location

Survey location was determined on the basis of the GPS coordinates captured when the adult roster was completed, prior to the selection of the respondent. These coordinates were used as they corresponded to the respondent's actual place of residence, and generally had good accuracy. These coordinates were used in the generation and validation of all spatial variables (i.e. municipality_coded, planning_region_code, ward_code, ea_code, etc.) included in the dataset.

3.1.2 Paradata variables included with the dataset

The paradata variables included in the QoL 7 (2023/24) dataset are listed and described in the table below.

Table 3: Paradata included with the QoL 7 (2023/24) dataset

Variable name	Description
interview_date	Full calendar date on which the interview was conducted.
District_municipality	Numerically coded variable providing the metropolitan or district municipality in which the interview was conducted, generated using the 'municipality_coded' variable.
municipality_coded	Numerically coded variable providing the metropolitan or local municipality in which the interview was conducted, generated using the geo-coordinates.
Planning_region	Text variable. For interviews in metropolitan municipalities, this variable provides the municipal planning region, generated using interview geo-coordinates. For interviews in local municipalities, the local municipality name is provided.
Planning_region_code	Text variable providing planning region or local municipality codes.

Variable name	Description
ward_code	Ward in which the interview was conducted, recorded during data collection and subsequently validated using interview geo-coordinates.
sp_name	Sub-place names are a suburb or townships in which the interview was conducted, generated using interview geo-coordinates.
ea_code	Enumeration area in which the interview was conducted, generated using interview geo-coordinates.
adult_count	Number of adult residents listed during the household listing prior to random selection of the respondent.
dur_mins	Duration of the fieldworker-administered questionnaire component in minutes, as recorded by the data collection system.
interview_lang	Interview language selected by the respondent, as recorded by the fieldworker at the beginning of the interview.
sc_consent	Whether the respondent consented to the self-complete questionnaire component.

3.2 Missing information

Standard codes are used in the dataset to represent information that is missing for various reasons. These are detailed in the table below.

Table 4: Standard codes of missing data

−1	Data missing due to a valid skip pattern (i.e. the respondent was not asked the question because it was skipped and not applicable to them)
−3	Data missing due to a fieldwork error (i.e. the question was not asked of the respondent but should have been, or a response was not appropriately recorded)

The following variables have data missing due to a fieldwork error:

- Q1.2d: Households sharing the room (n=3)
- Q1.7: Water from a rainwater tank (n=1)
- Q3.4: Year moved to Gauteng (n=7)
- Q5.1: Purpose of trip (n=1)

In section 15 (self-complete) of the questionnaire, all questions had a 'Prefer not to answer' response option. Coding of these responses is question-specific, and there is no standardised code. These responses have not been stipulated as missing in the SPSS dataset.

3.3 Standard response options

Default coding for questions with standard response options are listed below. However, the user should be guided firstly by the codes provided in the questionnaire and the labelling within the dataset if this differs from the coding detailed below.

3.3.1 Main questionnaire

Table 5: Default coding for Yes-No questions

1	Yes
0	No

Table 6: Default coding for Yes-No questions with 'Don't know' option

1	Yes
0	No
2	Don't know

Table 7: Default coding for Yes-No questions with 'three' options

1	Yes, in past 12 months
0	No
2	Yes, prior 12 months

Table 8: Default coding for satisfaction-scale questions

1	Very satisfied
2	Satisfied
3	Neither satisfied nor dissatisfied
4	Dissatisfied
5	Very dissatisfied

Table 9: Default coding for agreement-scale questions

1	Strongly agree
2	Agree
3	Neither agree nor disagree
4	Disagree
5	Strongly disagree

Table 10: Default coding for Gauteng local municipalities

1	City of Ekurhuleni
2	City of Johannesburg
3	City of Tshwane
4	Emfuleni
5	Lesedi
6	Midvaal
7	Merafong
8	Mogale City
9	Rand West

Table 11: Default coding for safety scale questions

1	Very safe
2	Fairly safe
3	Neither safe nor unsafe
4	Bit unsafe
5	Very unsafe

For Yes-No list questions, the dataset includes a variable for each response option. The standardised coding for these variables is shown in Table 12.

Table 12: Default coding for each response in multi-select or Yes-No list questions

1	Option selected by respondent
0	Option not selected by respondent

3.3.2 Self-complete questionnaire component

Two additional sets of standardised codes were used for the self-complete questionnaire component. These are presented in Tables 13 and 14. The Yes-No questions in the self-complete questionnaire component are coded in the same way as those in the fieldworker-administered component.

Table 13: Default coding for Yes-No questions with 'five' options (self-complete section only)

1	No, never
2	Yes, at home only
3	Yes, at school only
4	Yes, both at home and school
5	Prefer not to answer

Table 14: Default coding for Yes-No questions with 'Prefer not to answer' option (self-complete section only)

1	Yes
0	No
2	Prefer not to answer

4. RECODES AND DERIVED VARIABLES

The dataset includes a number of recodes and derived variables that have been used by the GCRO for analysis and may be of use to other data users. This section provides detail and coding for all recodes and derived variables included in the dataset. In all instances, the original variable is included as well should data users prefer to use the original variable or to generate their own recodes. There is one exception to this, which is the age variable. Due to the need to preserve anonymity, we have not been able to include the original age variable in the open public dataset but do provide a version top-coded to 80 years old, along with a fairly fine-grained recode, described in sub-section 4.1.11. The original age variable is available in the restricted access version of the dataset and should be requested from DataFirst with a motivation.

4.1 Data recodes

We provide a number of recodes within the dataset. Many of these are simply to provide more useful analytical categories, while others address concerns with particular variables. Most recodes contain 'recode' in the variable name and variable label, and details are provided below.

4.1.1 Interview duration

The variable 'dur_mins_recode' provides a categorical recode for interview durations in minutes, derived from 'dur_mins'. Categories are provided in the table below.

Table 15: Coding of 'dur_mins_recode' variable

Value	Label
1	15–20 minutes
2	21–30 minutes
3	31–40 minutes
4	41–50 minutes
5	51–60 minutes
6	61–90 minutes
7	More than 90 minutes

4.1.2 Month and year of interview

The variable 'date_month' is a categorical variable providing the month and year in which the interview was conducted, derived from 'interview_date'.

Table 16: **Coding of 'date_month' variable**

Value	Label
1	Aug 2023
2	Sep 2023
3	Oct 2023
4	Nov 2023
5	Dec 2023
6	Jan 2024
7	Feb 2024
8	Mar 2024
9	Apr 2024

4.1.3 Dwelling type (A3)

The variable 'a3_dwelling_type_recode' simplifies the many categories in 'a3_dwelling_type' into three main categories: 'Formal', 'Informal' and 'Other'.

Table 17: Recoding of 'a3_dwelling_type' into 'a3_dwelling_type_recode'

Original: a3_dwelling_type		a3_dwelling_type_recode	
Value	Label	Value	Label
1	House, brick or concrete structure on a separate stand	1	Formal
2	Traditional dwelling, hut or structure made of traditional materials	3	Other
3	Flat or apartment in a block of flats	1	Formal
4	Cluster house in a complex	1	Formal
5	Townhouse (semi-detached house in a complex)	1	Formal
6	Semi-detached house not in a complex	1	Formal
7	House, flat or room separate from main dwelling in backyard	1	Formal
8	Informal dwelling or shack in backyard	2	Informal
9	Informal dwelling NOT in backyard, e.g. in informal squatter settlement or on a farm	2	Informal

Original: a3_dwelling_type		a3_dwelling_type_recode	
10	Room or flat which is part of main dwelling or property	1	Formal
11	Caravan or tent	3	Other
12	Unit in a retirement home or barracks, etc.	1	Formal
13	Hostel	3	Other
14	Other (specify)	3	Other

4.1.4 Place of birth (Q3.1)

The variable 'q3_1_birth_prov_recode' simplifies the responses in 'q3_1_birth_prov' into three categories: 'Born in Gauteng', 'Born in another province in South Africa' and 'Born in another country'.

Table 18: Recoding of 'q3_1_birth_prov' into 'q3_1_birth_prov_recode'

Original: q3_1_birth_prov		q3_1_birth_prov_recode	
Value	Label	Value	Label
1	Gauteng	1	Born in Gauteng
2	Eastern Cape	2	Born in another province in South Africa
3	Free State	2	Born in another province in South Africa
4	KwaZulu-Natal	2	Born in another province in South Africa
5	Limpopo	2	Born in another province in South Africa
6	Mpumalanga	2	Born in another province in South Africa
7	Northern Cape	2	Born in another province in South Africa
8	North West	2	Born in another province in South Africa
9	Western Cape	2	Born in another province in South Africa
10	Country outside of South Africa	3	Born in another country

4.1.5 Timing of respondent's move to Gauteng

The variable 'q3_4_year_gp_recode' is a categorical variable indicating the duration of time since the respondent moved to Gauteng (if applicable). This was calculated first by using the year in which the interview was conducted together with the year in which the respondent moved to Gauteng to determine how many years ago the respondent had moved to the province, and then allocating the appropriate code as shown in the table below.

Table 19: Coding of 'q3_4_year_gp_recode'

Value	Label
1	In the last year
2	2–3 years ago
3	4–5 years ago
4	6–10 years ago
5	More than 10 years ago

4.1.6 Time of departure for most frequent trip

The categorical variable 'q5_5_depart_time_recode' provides departure time for the respondent's most frequent trip in hourly intervals. The variable is calculated from the time of departure as provided in 'q5_5_depart_time'. The coding for the hourly categories is provided below.

Table 20: Coding of 'q5_5_depart_time_recode'

Value	Label	Value	Label
1	00:00–00:59	2	01:00–01:59
3	02:00–02:59	4	03:00–03:59
5	04:00–04:59	6	05:00–05:59
7	06:00–06:59	8	07:00–07:59
9	08:00–08:59	10	09:00–09:59
11	10:00–10:59	12	11:00–11:59
13	12:00–12:59	14	13:00–13:59
15	14:00–14:59	16	15:00–15:59
17	16:00–16:59	18	17:00–17:59
19	18:00–18:59	20	19:00–19:59
21	20:00–20:59	22	21:00–21:59
23	22:00–22:59		

4.1.7 Travel duration for most frequent trip (Q5.6)

The categorical variable 'q5_6_time_destination_recode' provides the duration of the respondent's most frequent trip in 15 minute intervals. This variable is calculated based on the travel duration in minutes as specified by the respondent in the variable 'q5_6_time_destination'. Coding is provided in the table below.

Table 21: Coding of 'q5_6_time_destination_recode'

Value	Label
1	0–15 minutes
2	16–30 minutes
3	31–45 minutes
4	46–60 minutes
5	61–75 minutes
6	75–90 minutes
7	More than 90 minutes

4.1.8 Walking time to nearest public transport (Q5.11)

The categorical variables 'q5_11_pub_transport_prox_recode' provides the walking time to the respondent's nearest public transport access point in 10 minute intervals. This variable is calculated based on the walking time in minutes as specified by the respondent in the variable 'q5_11_pub_transport_prox'. Coding is provided in the table below.

Table 22: Coding of 'q5_11_pub_transport_prox_recode'

Value	Label
1	0–10 minutes
2	11–20 minutes
3	21–30 minutes
4	31–40 minutes
5	More than 40 minutes

4.1.9 Not enough money to feed children in household (Q6.5)

The variable 'q6_5_feed_children' has three response options: 'Yes', 'No' and 'There are no children in this household'. In some instances, a binary variable offering only 'Yes' and 'No' is easier to work with. For this reason, we include with the dataset the variable 'q6_5_feed_children_recode', coded with 'Yes' and 'No' responses, and households without children are set to missing values. Analysts should consider which variable is more appropriate for their purposes and ensure that they are interpreting the variable accurately. Details are provided in the table below.

Table 23: Coding of 'q6_5_feed_children_recode'

Original: q6_5_feed_children		q6_5_feed_children_recode	
Value	Label	Value	Label
0	No	0	No
1	Yes	1	Yes
2	There are no children in this household	–1	Missing value

4.1.10 Highest level of education attained (Q14.1)

The categorical variable 'q14_1_education_recode' reduces the responses regarding highest level of education attained, as captured in 'q14_1_education', into six categories. Coding is provided in the table below.

Table 24: Recoding of 'q14_1_education' into 'q14_1_education_recode'

Original: q14_1_education		q14_1_education_recode	
Value	Label	Value	Label
1	No education	1	No education
2	Grade 0 or Grade R	2	Primary only
3	Grade 1 or Sub A	2	Primary only
4	Grade 2 or Sub B	2	Primary only
5	Grade 3, Std 1	2	Primary only
6	Grade 4, Std 2	2	Primary only
7	Grade 5, Std 3 or ABET 1	2	Primary only
8	Grade 6, Std 4 or ABET 2	2	Primary only
9	Grade 7, Std 5	2	Primary only
10	Grade 8, Std 6, Form I or ABET 3	3	Secondary incomplete
11	Grade 9, Std 7, Form II, NQF 1 or ABET 4	3	Secondary incomplete

Original: q14_1_education		q14_1_education_recode	
Value	Label	Value	Label
12	Grade 10, Std 8, Form III, National Trade Certificate 1	3	Secondary incomplete
13	Grade 11, Std 9 or Form IV	3	Secondary incomplete
14	Grade 12, Std 10, Matric	4	Matric
15	A certificate from a college, technikon or university	5	More
16	A diploma from a college, technikon or university	5	More
17	Technikon or university degree	5	More
18	Postgraduate degree – e.g. Hons, MA, PhD	5	More
19	Unspecified	6	Unspecified

4.1.11 Respondent age (Q14.2)

As previously indicated, the raw respondent age variable, 'q14_2_age', is not included in the open public dataset due to the need to protect respondent anonymity. We do, however, include a version of the age variable, 'q14_2_age_topcode', which provides exact age (in years) for all respondents aged up to 79, and then codes all older respondents as 80. In addition, we provide a categorical variable, 'q14_2_age_recode' which converts the original age variable into 11 categories, as detailed in the table below. The original age variable is available in the restricted access version of the dataset and should be requested from DataFirst with a motivation.

Table 25: Coding of 'q14_2_age_recode'

Value	Label
1	18–19
2	20–24
3	25–29
4	30–34
5	35–39
6	40–44
7	45–49
8	50–54
9	55–59
10	60–64
11	65+

4.1.12 Number of household residents (Q14.5)

The variable 'q14_5_people_recode' provides a seven-category alternative to 'q14_5_people'. The recode retains the exact number of household residents for households with up to 6 residents and collapses all larger households into a single '7+' category, as shown in the table below.

Table 26: Coding of 'q14_5_people_recode'

Value	Label
1	1
2	2
3	3
4	4
5	5
6	6
7	7+

4.1.13 Number of household residents under 18 years of age (Q14.6)

The variable 'q14_6_under18_recode' provides a five category alternative to 'q14_6_under18'. The recode retains the exact number of household residents under 18 for households with up to three residents under 18 and collapses all households with larger numbers of residents under 18 into a single '4+' category, as shown in the table below.

Table 27: Coding of 'q14_6_under18_recode'

Value	Label
0	0
1	1
2	2
3	3
4	4+

4.1.14 Number of household residents aged 65 years or older (Q14.7)

The variable 'q14_7_65plus_recode' provides a four category alternative to 'q14_7_65plus'. The recode retains the exact number of household residents aged 65 years or older for households with up to two of these residents and collapses all households with larger numbers of residents 65 or above into a single '3+' category, as shown in the table below.

Table 28: Coding of 'q14_7_65plus_recode'

Value	Label
0	0
1	1
2	2
3	3+

4.1.15 Household income (Q15.3)

The categorical variable 'q15_3_income_recode' reduces the responses regarding monthly household income, as captured in 'q15_3_income', into seven categories. Coding is provided in the table below. Responses of 'no income', 'don't know' or 'prefer not to answer' were selected for 4 243 individuals. In the original variable, these responses each have a unique code and are not set as missing. In the recode, all are coded as 7 and set to missing. The reason for including those who selected 'no income' in this category was that in a large proportion of cases, other available data from the same respondents suggested that this was not a reliable response.

Table 29: Recoding of 'q15_3_income' into 'q15_3_income_recode'

q15_3income		q15_3_income_recode	
Value	Label	Value	Label
1	R1–R400	1	R1–R800
2	R401–R800	1	R1–R800
3	R801–R1 600	2	R801–R3 200
4	R1 601–R3 200	2	R801–R3 200
5	R3 201–R6 400	3	R3 201–R12 800
6	R6 401–R12 800	3	R3 201–R12 800
7	R12 801–R19 200	4	R12 801–R25 600
8	R19 201–R 25 600	4	R12 801–R25 600
9	R25 601–R38 400	5	R25 601–R51 200

q15_3income		q15_3_income_recode	
Value	Label	Value	Label
10	R38 401–R51 200	5	R25 601–R51 200
11	R51 201–R76 800	6	R51 201 and more
12	R76 801–R102 400	6	R51 201 and more
13	R102 401 -R153 600	6	R51 201 and more
14	R153 601–R204 800	6	R51 201 and more
15	R204 801–R500 000	6	R51 201 and more
16	More than R500 000	6	R51 201 and more
17	No income	7	No income/Prefer not to answer/Don't know
18	Prefer not to answer	7	No income/Prefer not to answer/Don't know
19	Don't know	7	No income/Prefer not to answer/Don't know

4.2 Derived variables included with the dataset

4.2.1 Access to any form of electricity (Q1.12)

Question 1.12, which asks respondents about all types of electricity supply that they use, was asked as a Yes-No list. This results in a series of 12 variables, each one representing a particular type of electricity supply, and coded 1 if the respondent said 'Yes' and 0 if 'No'. A subset of the variables (q1_12_postpaid, q1_12_8_dont_know and q 1_12_9_none) also have values missing due to skips, as detailed in sub-section 2.9.1.

We use the data from these 12 variables to generate the variable 'Any_electricity', coded 1 if the respondent has access to some of electricity, 0 if not and 2 if they indicated that they did not know. We categorised respondents as having access to electricity if they had indicated that they used any of the forms of electricity supply asked about. If they did not report using any of these forms of electricity supply, they were categorised as not having access to electricity. Details are provided in the table below.

We note that there are several different ways an analyst might approach creating this variable, and we encourage analysts to assess whether our derived variable is appropriate for their intended use. In particular, some analysts may prefer to make use of the variable q1_12_9_none in order to identify only those respondents who stated that they did not make use of electricity at all. Please see additional information on response patterns to this question, which may inform analytical decisions, in section 5.1 of this report.

Table 30: Coding of 'Any_electricity'

Value	Label	Definition
0	No form of electricity	All other respondents
1	Some form of electricity	q1_12_1_prepaid=1 OR q1_12_2_postpaid=1 OR q1_12_4_generator=1 OR q1_12_5_neighbour=1 OR q1_12_6_car_battery=1 OR q1_12_7_elsewhere=1 OR q1_12_10_pv_panels=1 OR q1_12_11_inverter=1 OR q1_12_12_wind_power=1
2	Don't know	q1_12_8_dont_know

4.2.2 Metered electricity connection (Q1.12)

We also used the data from Q1.12 to generate a variable which indicates whether or not the respondent reported using a metered electricity supply, whether prepaid or postpaid, 'Metered_connection'. The coding is detailed in the table below. Please note that in this variable we make use of a 'not applicable/unknown' category for those respondents who report not knowing the types of electricity supply used as well as those who report no access to electricity. In the SPSS version, the 'not applicable/unknown' category is stipulated as missing. Again, we encourage analysts to assess whether this approach is best suited for their particular purposes.

Table 31: Coding of 'Metered_connection'

Value	Label	Definition
0	Not a metered connection	All respondents not falling into the two categories below
1	Has a metered connection	q1_12_1_prepaid=1 OR q1_12_2_postpaid=1
2	Not applicable/unknown	q1_12_8_dont_know=1 OR q1_12_9_none=1

4.2.3 Currently generating own electricity (Q1.12)

A third derived variable drawing on Q1.12 is 'Generating_electricity', which indicates whether the respondent is already generating energy through either solar, wind or PV power, or through petrol or diesel generators. The coding is shown in the table below. For this variable, we also make use of a 'Not applicable/unknown' category for those who are not sure of their electricity source or do not have access to electricity. In the SPSS version, the 'Not applicable/unknown' category is stipulated as missing.

Table 32: Coding of 'Generating_electricity'

Value	Label	Definition
0	Not generating electricity	All respondents not falling into the three categories below
1	Generating some electricity	q1_12_4_generator=1 OR q1_12_10_pv_panels=1 OR q1_12_12_wind_power=1
2	Not applicable/unknown	q1_12_8_dont_know=1 OR q1_12_9_none=1

4.2.4 Considering investing in alternative sources of energy (Q1.12b)

Respondents were asked a series of questions (as Yes-No) about their considerations about investing (or further investing) in alternative sources of energy. We include with the dataset the variable 'Investment_alt_energy', which is coded 1 if the respondent reported that their household considered investing or further investing ('Yes') in any of the five alternative energy sources. It is coded 0 if the respondent did not report considering investing or further investing in any of the alternative energy sources. Please note that individuals who reported in Q1_12_9_none that they had no access to electricity were also asked these questions. The coding is provided below.

Table 33: Coding of 'plan_to_invest_alt_energy'

Value	Label	Definition
0	Do not consider investing in alternative energy	All respondents not meeting the criteria below
1	Plan to invest in alternative energy	q1_12b_1_invest_solar=1 OR q1_12b_2_invest_pv_panels=1 OR q1_12b_3_invest_inverter=1 OR q1_12b_4_invest_generator=1 OR q1_12b_5_invest_windpower=1

4.2.5 Purpose of most frequent trip (Q5.1 & Q5.3)

Respondents were asked about the purpose of the most frequent trip they make from home in Q5.1. These responses are recorded in the variable 'q5_1_frequent_trip'. However, as detailed in sub-section 2.9.1, those who indicated that they never went anywhere in this question were asked in Q5.2 to verify this. A small number of respondents (n=19) indicated at this point that they did go somewhere, and the purpose of their trip is recorded in 'q5_3_trip'. For ease of analysis, we have created the variable 'q5_1_3_frequent_trip_cons', which uses 'q5_1_frequent_trip' as a base and then updates the information for the 19 respondents who provided a trip purpose in 'q5_3_trip'. In other words, Q5.1 and Q5.3 are combined to include those who first said they don't make any trips. This is considered the final trip purpose variable, and we recommend using 'q5_1_3_frequent_trip_cons' for any analysis.

4.2.6 Participation in organised social activities (Q12.1)

The question about whether the respondent participated in any of a range of organised social activities (Q12.1) was asked as a Yes-No list. Consequently, the dataset includes a binary variable for each activity type, coded '1' if the respondent reported participating in the activity and '0' if they did not. The variable 'Social participation' combines the data from these six variables into a single indicator of whether the respondent participated in any of these forms of social activity in the past year.

Table 34: Coding of 'Social_participation'

Value	Label	Definition
0	No participation	All respondents not meeting the criteria below
1	Some participation	q12_1_1_church = 1 OR q12_1_2_social = 1 OR q12_1_3_stokvel = 1 OR q12_1_4_community = 1 OR q12_1_5_political = 1 OR q12_1_6_other = 1

4.2.7 Involvement in participatory governance activities (Q12.2)

The question about whether the respondent or a household member participated in any of a range of participatory governance activities (Q12.2) was asked as a Yes-No list. Consequently, the dataset includes a binary variable for each activity type, coded '1' if the respondent reported participation in the activity and '0' if they did not. The variable 'Political_participation' combines the data from these seven variables into a single indicator of whether the respondent or other household member participated in any form of participatory governance activity in the past year.

Table 35: Coding of 'Political_participation'

Value	Label	Definition
0	No participation	All respondents not meeting the criteria below
1	Some participation	q12_2_1_ward = 1 OR q12_2_2_street = 1 OR q12_2_3_cdf = 1 OR q12_2_4_idp = 1 OR q12_2_5_mayor = 1 OR q12_2_6_sgb = 1 OR q12_2_7_cpf = 1

4.2.8 PHQ-2 score (Q13.9 & Q13.10)

The 2023/24 survey included the two four-point scale items that comprise the Patient Health Questionnaire–2 (PHQ-2) score, a short screening tool for possible depressive symptoms (Kroenke et al., 2003). The two items are the frequency of loss of interest or pleasure in things over the past two weeks ('q13_9_pleasure') and the frequency of feeling depressed over the past two weeks ('q13_10_depressed'). Each is coded on a scale running from 'Not at all' (coded 1) to 'Nearly every day' (coded 4).

To calculate the PHQ-2 score, responses to Q13.9 and 13.10 are added and two (2) is subtracted from the total sum to a composite variable (PHQ2_score), with scores ranging from zero (0) to six (6). Scores below three (<3) indicate that the respondent is not at high risk of depression, while higher scores (>=3) indicate the respondent is at high risk. This is not a measure for clinical depression, but a brief screening tool for the risk of depressive symptoms. Researchers/analysts are advised to consult the relevant literature for guidance on the use of this score.

4.2.9 PHQ2_score_high (Q13.9 & Q13.10)

A widely used approach to the interpretation of PHQ-2 scores is to use a score of three (3) or higher as an indication that an individual is at high risk of depression. The variable 'PHQ2_score_high' is calculated on this basis, and is coded 1 if a respondent is at high risk of depression and 0 if a respondent is not. The coding is detailed in the table below.

Table 36: Coding of 'PHQ2_score_high'

Value	Label	Definition
0	Not at high risk of depression	PHQ2_score < 3
1	High risk of depression	PHQ2_score >= 3

4.2.10 QoL Index variables

The QoL 7 (2023/24) dataset includes key variables related to the GCRO's QoL Index. This section provides only a brief overview of the variables included in the dataset. For details on the derivation and calculation of the Index, see Naidoo and de Kadt (2024).

Dimension scores: The dataset includes the score for each of the seven dimensions feeding into the composite QoL Index. These variables are each scaled to run from 0 to 10, with lower scores indicating lower levels of well-being in that domain, and higher scores indicating higher levels of well-being. Each variable is suitable for use on its own as a measure of well-being in that respective dimension. The dimension score variables are as follows:

- Services ('F1servic')
- Socio-economic status ('F2soclas')
- Government satisfaction ('F3govsat')
- Life satisfaction ('F4lifsat')
- Health ('F5health')
- Safety ('F6safety')
- Participation ('F7partic')

Composite QoL Index score: The overall GCRO QoL Index score for each respondent is available in 'QoLIndex_Data_Driven'. This variable is scaled to run from 0 to 100, with lower scores indicative of a poorer quality of life and higher scores indicative of a higher quality of life.

5. ADDITIONAL INFORMATION ON SELECTED VARIABLES AND SPATIAL AREAS

This section provides some additional information on a number of questions and variables which may be useful to a data user. This covers some implementation challenges impacting a small number of variables, as well as other information relevant to the interpretation of certain variables. This information is drawn from the pilot process, the main data collection process, various debrief activities and the analysis already undertaken by the GCRO.

Where implementation issues have resulted in missing information or data that is challenging to interpret, we have not attempted to correct the data, unless otherwise specified, preferring to allow each analyst to make their own informed decision. However, in some instances we have provided recodes in addition to the original data. Where recodes are available, this is indicated, and further information is available in sub-section 4.1.

5.1 Types of electricity used (Q1.12)

Some unexpected response patterns to Q1.12 regarding types of electricity supply makes interpretation of the findings slightly complex. While most response patterns across the 12 items in this Yes-No list make sense, there are 533 respondents (3.9% of the sample) who answered 'No' to all 12 options, meaning that we do not actually know whether or not they have electricity. These respondents answered 'No' to all possible electricity sources, including 'Other', but also answered 'No' when asked if they have no electricity at all. Discussions with fieldworkers regarding this pattern suggested that some respondents were extremely unwilling to provide any information regarding electricity supply or use. This may relate to use of illegal connections and concern about possible reprisals.

A further 192 respondents indicated that they didn't know what type of electricity supply the household had. While in some instances this is likely to be an accurate response, there is also a possibility that it was used in some cases due to a reluctance to disclose illegal connections.

5.2 Presence of children in household (Q6.5 and Q14.6)

We collected information on the presence of children in the household in two different parts of the questionnaire. In Q6.5, we asked whether there had been insufficient money to feed children in the household in the past 12 months and included a response option of 'There are no children in this household'. In Q14.6, we asked the respondent to provide us with the number of household residents under the age of 18 years.

For a number of respondents, there was an inconsistency between the responses to these two questions. For 6 441 respondents who reported no residents under the age of 18, 113 reported difficulty feeding children in Q6.5, and a further 1 009 reported no difficulty feeding children in Q6.5.

It was not possible to check all responses through callbacks, but feedback from field notes and fieldworkers, along with a small number of callbacks, suggested a number of explanations for various response patterns. These are:

- We believe that changes in household configurations over the preceding 12 months are likely to explain a proportion of the inconsistent responses, and particularly those where respondents reported difficulty in feeding children but no household residents under the age of 18. We believe that in many of the cases in which respondents report difficulty in feeding children but no residents under the age of 18, it is likely that children may have moved out of the household in question. In some instances, children might also have passed away and a small number might have only very recently turned 18.
- Where respondents reported no residents under the age of 18 but responded to Q6.5 indicating no difficulty in feeding children, we believe the majority of these cases relate to participants not being offered the full set of response options before responding. That is, we believe that many respondents without children immediately said 'No' when asked about difficulty in feeding children and fieldworkers did not check with the respondent as to whether there were any children.
- We also believe that not all respondents understood the term 'children' in Q6.5 to include all individuals under the age of 18. In particular, feedback indicated that some respondents did not understand the term 'children' to include babies, and that others did not understand it to include adolescents, and particularly older adolescents.

5.3 Medical conditions of household members (Q13.11)

Question 13.11, implemented as an extended Yes-No list, asked respondents whether they or another household member had had a range of medical conditions during or prior to the past year (coded as 0=No, 1=Yes, in past 12 months and 2=Yes, prior to past 12 months). During training, fieldworkers were told that the condition should have been diagnosed by a medical practitioner for it to be recorded as Yes, either in the past year or prior. It should be noted that some respondents were more aware than others of the medical conditions of household members, and this should also be considered when analysing this data. Where a respondent was unsure, or did not want to respond to a particular item, this was recorded as 'No'. Please be aware that these variables should be used with care when attempting to assess levels of particular medical conditions in an area, as they do not provide information on the number of household members experiencing a particular condition. Furthermore, it is also important to note that the data from this question cannot be used to measure disease prevalence.

5.4 COVID-19 and vaccination (Q13.15-Q13.26)

In QoL 6 (2020/21), several questions about COVID-19 were introduced to assess how the pandemic might have affected respondents' socio-economic and health circumstances. In QoL 7 (2023/24), new questions were added to explore post-pandemic distress and attitudes toward vaccines. Most of these questions were spontaneous mentions and asked of all respondents, except where a question was not applicable based on previous answers. For example, Q13.15 was asked of everyone, but only those who answered 'I thought I had COVID-19 but did not seek medical care' were asked Q13.16. Fieldworkers were specifically trained to ensure Q13.15 was answered accurately by probing whether the respondent had tested positive rather than just how they managed their health after they thought they had COVID-19, as subsequent questions depended on the response provided in Q13.15.

Additionally, probing was done in Q13.20, where respondents were asked whether it was easy or difficult to get their most recent vaccination. Since there was no option for 'it was difficult' on the questionnaire for Q13.20, respondents who initially reported difficulty as a spontaneous response were asked to explain why it was difficult. This allowed the fieldworker to accurately select the appropriate response according to the survey questionnaire.

Question 13.26 asked respondents whether they had ensured that all dependents or children in their household for whom they were responsible had received all necessary vaccinations. Although Q6.5 included an option indicating 'there are no children in your household', Q13.26 was asked of everyone. Discrepancies may arise regarding the presence of children in the household. For instance, if a respondent selected 'No children' in Q6.5, the number of respondents indicating 'I am not responsible for children' in Q13.26 should be higher than the frequency reported in Q6.5. Data users are advised to use Q13.26 with caution and careful scrutiny.

5.5 Household income (Q15.3)

In all previous QoL survey iterations, except QoL 6 (2020/21), fieldworkers asked respondents to provide their household income as part of the fieldworker-administered questionnaire. Typically, around a third of respondents refused to answer this question due to its sensitivity.

In QoL 6 (2020/21), which included a self-complete questionnaire component, we decided to include the household income question in this module to assess whether it might improve response rates. However, we found that 14% of the sample chose not to complete this section, preventing us from linking their income levels to other questions in the fieldworker-administered survey.

In QoL 7 (2023/24), we moved Q15.3 back to the main questionnaire, hoping respondents would be more willing to share their household incomes. Similar to previous iterations, a quarter of respondents (n=3 414) selected 'prefer not to answer', 803 respondents (6%) selected 'don't know' and 26 respondents (0.2%) selected 'no income'. For this reason, we recommend caution in working with these responses and, in our recode (q15_3_income_recode), we have adjusted these to one combined category that should be treated as missing. Finally, due to the change in how this question was asked relative to QoL 6 (2020/21), we encourage analysts undertaking longitudinal comparisons to reflect on whether changes in response patterns might impact on their analyses.

References

de Fortier, E. and Loots, H. (2024). *Fieldwork report: Quality of Life Survey 7 (2023/24)*. Johannesburg: Gauteng City-Region Observatory. https://doi.org/10.36634/EOSA3094

Naidoo, L., Hamann, C. and Naidoo, Y. (2024). *Sampling report: GCRO Quality of Life Survey 7 (2023/24)*. Johannesburg: Gauteng City-Region Observatory. https://doi.org/10.36634/EOSA3096

Kroenke, K., Spitzer, R.L. and Williams, J.B. (2003). 'The Patient Health Questionnaire-2: Validity of a two-item depression screener'. *Medical Care*, 1284–1292. https://doi.org/10.1097/01.mlr.0000093487.78664.3c

Naidoo, Y. and de Kadt, J. (2024). *Quality of Life Index methodology: Quality of Life Survey 7 (2023/24)*. Johannesburg: Gauteng City-Region Observatory (GCRO). https://doi.org/10.36634/TBPJ1907

Neethling, A. (2024). *Weighting report: Quality of Life 7 Survey (2023/24)*. Johannesburg: Gauteng City-Region Observatory. https://doi.org/10.36634/EOSA3095

Orkin, M. (2020). *Technical review of the GCRO QoL Surveys: Synthesis report*. Johannesburg: Gauteng City-Region Observatory.

Statistics South Africa. (2012). *South African Standard Classification of Occupations (SASCO) 2012*. Pretoria: Statistics South Africa.

About the authors

Sthembiso Pollen Mkhize | https://orcid.org/0000-0003-4360-3234

Sthembiso Pollen Mkhize was a Junior Researcher at the GCRO, where he led the Queering Social Survey Research project. He was also a valuable member of the team responsible for overseeing the flagship Quality of Life Survey from 2019 to 2024. Mkhize obtained his Master's degree in Population Studies from the University of KwaZulu-Natal and is currently pursuing his PhD in Human Geography at the University of Bristol. His passion revolves around various critical areas related to social justice, inclusion/exclusion, youth health, gender and sexuality, and quality of life and well-being.

Pedzisai Ndagurwa | https://orcid.org/0000-0001-5628-4901

At the time of writing of this Data Brief, Pedzisai Ndagurwa was a Senior Researcher at the GCRO, in that role between August 2022 and August 2023. At the GCRO, he served as the lead of the 'Understanding quality of life' research theme and led the Quality of Life 7 Survey (2023/24) implementation. Pedzisai holds a PhD from the University of the Witwatersrand in Demography and Population Studies. He also holds a Master's degree in Population Studies and a Bachelor's degree in Social Sciences (Industrial Psychology and Management) from the University of KwaZulu-Natal. Pedzisai is currently a Senior Lecturer in Demography and Population Studies at the University of the Witwatersrand.

Shamsunisaa Miles-Timotheus | https://orcid.org/0000-0003-1875-2465

At the time of writing this Data Brief, Shamsunisaa Miles-Timotheus was a GCRO intern working on the implementation of the Quality of Life 7 (2023/24) Survey. She is currently serving as a Junior Researcher for Quality of Life at the GCRO. Shamsunisaa obtained her Master's degree in Health Demography from the University of the Witwatersrand in 2021, and she also holds an Honours in Psychology and a Bachelor's degree in Sociology and Psychology, also from the University of the Witwatersrand.

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