



South African E-Cigarette Survey 2022

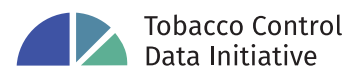
User Manual

May 2023

Authors

Kirsten van der Zee

Thomas van Huyssteen



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD



South African E-Cigarette Survey 2022

User Manual

Authors

Kirsten van der Zee¹

Thomas van Huyssteen²

1. Research Unit on the Economics of Excisable Products. School of Economics. University of Cape Town.
Kirsten.vanderzee@uct.ac.za

2. Corresponding author. Firdale Consulting. Thomas@firdaleconsulting.com

Table of Contents

1.	Introduction and Background	3
2.	Using this manual	3
	Citation guide	3
3.	Definitions	4
4.	Sampling, stratification and coverage	5
	Sample frame	5
	Stratification	6
5.	Weighting	11
6.	Questionnaire	11
7.	Data collection plan and timeline	12
7.1.	Initial timeline	12
7.2.	Piloting	12
7.3.	Data collection timeline for Phases 1 to 3	13
7.4.	Sample management between phases	14
8.	Final sample	15
9.	Caveats and problematic variables	16
9.1.	Initial routing question	16
9.2.	Changes to the questionnaire	16
9.3.	Date variable	18
10.	Quality control	19
11.	Ethics clearance	19

1. Introduction and Background

The Tobacco Control Data Initiative (TCDI) is implemented by Development Gateway: An IREX Venture, in partnership with the Research Unit on the Economics of Excisable Products (REEP) based at the University of Cape Town. The TCDI is funded by a grant from the Bill & Melinda Gates Foundation. The initiative aims to provide Sub-Saharan African governments with improved access to country-specific data to help them achieve better tobacco-control policy design and implementation.

The South African E-Cigarette Survey was conducted between January and September 2022 as part of the TCDI project. The aim of the survey was to provide data related to the use of electronic nicotine and non-nicotine delivery systems and heated tobacco products (collectively defined as ENDS, or “e-cigarettes” in the study), as well as of traditional cigarettes, amongst adults in urban South Africa.

The e-cigarette survey was designed to estimate the prevalence of e-cigarette use in urban South Africa, and to describe the demographics of e-cigarette users (age, gender, etc.). This study also sought to understand the interaction between cigarette smoking and e-cigarette use, at the individual level. Of particular interest was drawing a timeline of use for both cigarettes and e-cigarettes (including initiation and, where relevant, quitting) for each respondent, where dual use had occurred. TCDI hired Ask Afrika, a data collection company, to conduct the survey.

2. Using this manual

This manual is released together with the data from the South African E-Cigarette Survey 2022. The purpose of the manual is to give users of the e-cigarette survey data a thorough background understanding of the data-collection process. The manual outlines the data collection plan, including the sample design and stratification, the weighting of the final data, and how the data collection played out in the field. The manual also provides information on any changes made to the questionnaire during the data collection period, and any problematic variables and caveats that users should bear in mind when analysing the data.

Citation guide

The survey dataset should be cited as:

Tobacco Control Data Initiative. South African E-Cigarette Survey 2022 [dataset]. Version 1. Washington: TCDI [producer], 2023. Cape Town: DataFirst [distributor], 2023. DOI: <https://doi.org/10.25828/23YT-7V09>

The manual should be cited as:

Van der Zee, K., and van Huyssteen, T. 2023. South African E-Cigarette Survey 2022: User Manual. Washington: Tobacco Control Data Initiative.

3. Definitions

Because a key research objective of the survey was to understand the sequence of use between e-cigarettes and traditional cigarettes (including quitting), the questionnaire was designed to identify both current and past users of both products. The following table defines the terms used to describe the different types of users identified in the survey, for clarity both in the documentation that accompanies the data, as well as in the e-cigarette data.

Table 1: Definitions used in the e-cigarette survey

ENDS	Electronic nicotine and non-nicotine delivery system. Referred to interchangeably with “e-cigarettes” in this document and the dataset. The questionnaire refers to “e-cigarette, heated tobacco product, or other vape”, thus all of these groups are included when referred to in this and other project documents, and in the e-cigarette data.
Current ENDS user	Respondent who currently uses an e-cigarette at least once a week.
Current cigarette smoker	Respondent who has smoked at least 100 cigarettes in their lifetime, and currently smokes cigarettes daily, or at least once a week.
Past ENDS user	Respondent who is not a current e-cigarette user, but has used an e-cigarette at least weekly in a typical month.
Past cigarette smoker	Respondent who is not a current cigarette smoker, but who has smoked at least 100 cigarettes in their lifetime, and has smoked at least once a week, in a typical week.
ENDS experimenter	Respondent who has smoked at least one puff of an e-cigarette, but has never used an e-cigarette at least weekly, in a typical month.
Smoking experimenter	Respondent who has smoked at least one puff of a standard cigarette, but has smoked fewer than 100 cigarettes in their lifetime OR has never smoked cigarettes at least weekly.
Dual users	Combination of the above definitions using BOTH current and past use.
Cigarette user	Either current or past cigarette smoker.
E-cigarette user	Either current or past e-cigarette user.
Cigarette-only user	Either current or past cigarette smoker, who is not a current or past e-cigarette user.
E-cigarette only user	Either current or past e-cigarette user, who is not a current or past cigarette smoker.

4. Sampling, stratification and coverage

The data collection process was designed to be representative of adults (18 years and older) living in urban areas in South Africa. Sampling calculations suggested that we required 1250 interviews of e-cigarette users and 700 interviews of cigarette-only users, to answer our primary research question on sequence of use (i.e., what proportion of cigarette smokers used e-cigarettes before starting smoking, and what proportion of people who quit smoking cigarettes used e-cigarettes before quitting).

Prior evidence on e-cigarette use prevalence in South Africa (~3%) and telephone survey response rates suggested the enumerators should call 400 000 numbers to achieve 1250 completed surveys with e-cigarette users. Three different questionnaires were used for survey implementation: a long-form survey questionnaire for e-cigarette users, a long-form survey questionnaire for cigarette smokers, and a short-form survey questionnaire for all other respondents (non-users, and the cigarette-only users interviewed after the quota for this group was collected).

Sample frame

The sampling frame used for the survey was developed by the data collection company, Ask Afrika, in collaboration with Geoterraimage (GTI). The basis for the sampling frame was the sampling frame of Enumerator Areas (EAs) developed by Statistics SA (StatsSA) for the 1996 and 2001 censuses, updated annually using StatsSA's the mid-year population estimates.

Additionally GTI include in the sampling frame:

- Spatial data that reflect new building developments and urbanisation not identified by StatsSA which is integrated with the latest GTI Land Use dataset.
- The Residential Building Census (e.g., formal and informal dwelling structures) from GTI's New Developments© dataset

This updated sampling frame includes inter alia new developments, changes of EA type, and changes in administrative boundaries, such as municipality and ward boundaries.

The current EA sampling frame consists of 14 039 main places in South Africa. These main places are subdivided into 22 108 sub-places, which are further divided into 103 576 EAs.

The EA frame data were overlaid with Credit Bureau data, which consist of data on all individuals 18 years and older within South Africa who have a credit history with credit providers such as banks, insurance companies, or non-bank lenders, or with retailers of commodities such as clothing, furniture, or food (store cards). The credit bureau data includes records on incidental credit, such as DSTV accounts and phone contracts, as well as data on individuals who made inquiries about opening an account, even if the inquiry was unsuccessful or the offer of an account was not taken up.

The table below shows how the GTI EA frame represents the South African, population in total, as well the urban adult population (individuals aged 18 years and older) (adults), and the stratification variables used in the study.

Table 2: The sample frame by province, population group and Neighbourhood Lifestyle Index

	South Africa overall	Individuals aged 18+	Individuals aged 18+ in urban metro and non-metro areas
Province			
Eastern Cape	6 734 001	4 155 486	1 862 719
Free State	2 928 903	1 945 849	1 696 809
Gauteng	15 488 137	11 203 070	10 256 732
Kwa-Zulu Natal	11 531 627	7 222 294	4 012 691
Limpopo	5 852 552	3 557 508	655 274
Mpumalanga	4 679 787	3 063 360	1 768 110
North West	4 108 817	2 710 658	1 304 161
Northern Cape	1 292 787	853 975	645 680
Western Cape	7 005 740	4 994 461	4 446 090
Total	59 622 351	39 706 661	26 648 266
Population Group			
Black	48 153 728	31 162 934	18 939 977
Coloured	5 247 740	3 571 363	3 165 951
Indian	1 541 113	1 192 737	1 137 482
White	4 679 770	3 779 628	3 404 856
Total	59 622 351	39 706 662	26 648 266
Neighbourhood Lifestyle Index			
NLI1-3	45 367 598	29 169 334	16 869 533
NLI4-6	9 613 700	6 896 399	6 413 150
NLI7-10	4 600 739	3 605 972	3 365 583
Total	59 582 037	*39 671 705	26 648 266

Note: The Neighbourhood Lifestyle Index is an income segmentation variable, ranging from 1-10, where 10 indicates the highest-income households. The NLI is averaged over the EAs.

* The NLI total is less than the totals for population group and province; this is because some people live in EAs in which there is insufficient information to calculate NLI or in non-residential units, e.g. prisons.

Stratification

We applied a stratified random sample approach, by dividing the population into several mutually exclusive strata, to ensure that all important domains were represented in the sample and to ensure the most representative sample (the least variation and highest precision). The stratification variables used are as follows:

Explicit stratification variables

- Primary stratification: nine provinces and two geo-types (metro and urban non-metro);
- Secondary stratification: four population groups and three neighbourhood lifestyle index (NLI) segments (area income segments).

Implicit stratification variables

- District municipality, local municipality, main place, sub-place, and EA code.

The explicit stratification variables are used to ensure good coverage of the main demographics in SA and the best possible precision per stratum. The implicit stratification variables are used to improve the representativeness and coverage within each stratum. Within each of the explicit strata, the implicit stratification variables were used first to order all individuals in the strata (by the implicit variables), before systematically drawing the individuals for the sample. In total, 140 strata were created. Table 3 summarises the two primary stratification variables with each of the secondary stratification variables.

Table 3: Sample allocation for n=1250 e-cigarette interviews, by province, geo area, population group, and income group

Province	Geo area	Income group	Black	Coloured	Indian	White
Western Cape	Metro	Lower Income	32	27	1	1
Western Cape	Metro	Middle Income	10	22	4	11
Western Cape	Metro	Upper Income	2	4	1	16
Western Cape	Non-Metro Urban	Lower Income	11	16	1	2
Western Cape	Non-Metro Urban	Middle Income	3	9	1	9
Western Cape	Non-Metro Urban	Upper Income	0	1	0	3
Eastern Cape	Metro	Lower Income	31	11	1	3
Eastern Cape	Metro	Middle Income	10	7	2	9
Eastern Cape	Metro	Upper Income	1	0	0	4
Eastern Cape	Non-Metro Urban	Lower Income	17	6	0	1
Eastern Cape	Non-Metro Urban	Middle Income	7	3	1	4
Eastern Cape	Non-Metro Urban	Upper Income	2	0	0	1
Northern Cape	Non-Metro Urban	Lower Income	26	15	2	4
Northern Cape	Non-Metro Urban	Middle Income	3	13	0	6
Northern Cape	Non-Metro Urban	Upper Income	1	0	0	2
Free State	Metro	Lower Income	18	2	1	0
Free State	Metro	Middle Income	7	2	1	4
Free State	Metro	Upper Income	1	0	0	2
Free State	Non-Metro Urban	Lower Income	43	3	1	2
Free State	Non-Metro Urban	Middle Income	12	4	1	10
Free State	Non-Metro Urban	Upper Income	1	0	0	1
KwaZulu-Natal	Metro	Lower Income	43	0	8	1
KwaZulu-Natal	Metro	Middle Income	17	7	19	6

KwaZulu-Natal	Metro	Upper Income	3	0	3	6
KwaZulu-Natal	Non-Metro Urban	Lower Income	27	0	3	2
KwaZulu-Natal	Non-Metro Urban	Middle Income	13	3	7	6
KwaZulu-Natal	Non-Metro Urban	Upper Income	1	0	1	2
North West	Non-Metro Urban	Lower Income	47	2	2	6
North West	Non-Metro Urban	Middle Income	17	6	3	12
North West	Non-Metro Urban	Upper Income	2	0	0	4
Gauteng	Metro	Lower Income	104	9	3	3
Gauteng	Metro	Middle Income	42	12	11	18
Gauteng	Metro	Upper Income	16	2	5	25
Gauteng	Non-Metro Urban	Lower Income	15	1	0	1
Gauteng	Non-Metro Urban	Middle Income	7	1	1	5
Gauteng	Non-Metro Urban	Upper Income	1	0	0	2
Mpumalanga	Non-Metro Urban	Lower Income	57	1	3	4
Mpumalanga	Non-Metro Urban	Middle Income	24	3	3	10
Mpumalanga	Non-Metro Urban	Upper Income	5	0	1	7
Limpopo	Non-Metro Urban	Lower Income	31	0	1	2
Limpopo	Non-Metro Urban	Middle Income	18	3	3	7
Limpopo	Non-Metro Urban	Upper Income	5	0	0	3

The second step in the stratification process was to combine all the EAs per stratum. The sample frame has details of the population total, and age, gender, and race breakdown of the population per EA. Because of the large differences in the population sizes between the strata, a disproportionate sample was drawn to ensure smaller strata were boosted for more precision, and this was accounted for during the weighting phase.

Based on a previous study of people aged 15-years and older in urban areas of SA, the anticipated response rates for e-cigarette users was 3%. For the e-cigarette survey to yield a dataset with 1250 e-cigarette users, from an expected incidence of 3%, we would need to reach roughly 40 000 respondents ($40\,000 \times 3\% = 1200$). The distribution of these 40 000 sampling points is represented in the table below.

Table 4: Distribution of sampling points based on 3% incidence rate, for n=40 000, by province, geo area, population group, and income group

Province	Geo area	Income group	Black	Coloured	Indian	White
Western Cape	Metro	Lower Income	1024	864	32	32
Western Cape	Metro	Middle Income	320	704	128	352
Western Cape	Metro	Upper Income	64	128	32	512
Western Cape	Non-Metro Urban	Lower Income	352	512	32	64
Western Cape	Non-Metro Urban	Middle Income	96	288	32	288
Western Cape	Non-Metro Urban	Upper Income	0	32	0	96
Eastern Cape	Metro	Lower Income	992	352	32	96

Eastern Cape	Metro	Middle Income	320	224	64	288
Eastern Cape	Metro	Upper Income	32	0	0	128
Eastern Cape	Non-Metro Urban	Lower Income	544	192	0	32
Eastern Cape	Non-Metro Urban	Middle Income	224	96	32	128
Eastern Cape	Non-Metro Urban	Upper Income	64	0	0	32
Northern Cape	Non-Metro Urban	Lower Income	832	480	64	128
Northern Cape	Non-Metro Urban	Middle Income	96	416	0	192
Northern Cape	Non-Metro Urban	Upper Income	32	0	0	64
Free State	Metro	Lower Income	576	64	32	0
Free State	Metro	Middle Income	224	64	32	128
KwaZulu-Natal	Non-Metro Urban	Middle Income	416	96	224	192
KwaZulu-Natal	Non-Metro Urban	Upper Income	32	0	32	64
North West	Non-Metro Urban	Lower Income	1504	64	64	192
North West	Non-Metro Urban	Middle Income	544	192	96	384
North West	Non-Metro Urban	Upper Income	64	0	0	128
Gauteng	Metro	Lower Income	3328	288	96	96
Gauteng	Metro	Middle Income	1344	384	352	576
Gauteng	Metro	Upper Income	512	64	160	800
Gauteng	Non-Metro Urban	Lower Income	480	32	0	32
Gauteng	Non-Metro Urban	Middle Income	224	32	32	160
Gauteng	Non-Metro Urban	Upper Income	32	0	0	64
Mpumalanga	Non-Metro Urban	Lower Income	1824	32	96	128
Mpumalanga	Non-Metro Urban	Middle Income	768	96	96	320
Mpumalanga	Non-Metro Urban	Upper Income	160	0	32	224
Limpopo	Non-Metro Urban	Lower Income	992	0	32	64
Limpopo	Non-Metro Urban	Middle Income	576	96	96	224
Limpopo	Non-Metro Urban	Upper Income	160	0	0	96

The third step was to overlay these strata with the Credit Bureau data. The contact numbers listed in a stratum's EAs were ordered according to the implicit stratification variables and then systematically drawn to get a sample spread across the stratum, assuming a 1:10 contact rate (for every ten numbers drawn and contact attempted, one will result in contact with a respondent).

The number of credit bureau contact numbers per strata required to achieve the 1250 contacts with e-cigarette users (assuming a 1:10³ contact rate and 3% incidence) are shown in Table 5 (total of 400 000 contact numbers).

3. Initially it was assumed that a 1:20 contact rate would be achieved. However, after the first two phases of data collection, it was established that the contact rate was close to 1:10. See section 6.4: "Sample management between phases".

Table 5: Number of Credit Bureau contact numbers for n=400 000, by province, geo area, population group, and income group

Province	Geo area	Income group	Black	Coloured	Indian	White
Western Cape	Metro	Lower Income	10240	8640	320	320
Western Cape	Metro	Middle Income	3200	7040	1280	3520
Western Cape	Metro	Upper Income	640	1280	320	5120
Western Cape	Non-Metro Urban	Lower Income	3520	5120	320	640
Western Cape	Non-Metro Urban	Middle Income	960	2880	320	2880
Western Cape	Non-Metro Urban	Upper Income	0	320	0	960
Eastern Cape	Metro	Lower Income	9920	3520	320	960
Eastern Cape	Metro	Middle Income	3200	2240	640	2880
Eastern Cape	Metro	Upper Income	320	0	0	1280
Eastern Cape	Non-Metro Urban	Lower Income	5440	1920	0	320
Eastern Cape	Non-Metro Urban	Middle Income	2240	960	320	1280
Eastern Cape	Non-Metro Urban	Upper Income	640	0	0	320
Northern Cape	Non-Metro Urban	Lower Income	8320	4800	640	1280
Northern Cape	Non-Metro Urban	Middle Income	960	4160	0	1920
Northern Cape	Non-Metro Urban	Upper Income	320	0	0	640
Free State	Metro	Lower Income	5760	640	320	0
Free State	Metro	Middle Income	2240	640	320	1280
Free State	Metro	Upper Income	320	0	0	640
Free State	Non-Metro Urban	Lower Income	13760	960	320	640
Free State	Non-Metro Urban	Middle Income	3840	1280	320	3200
Free State	Non-Metro Urban	Upper Income	320	0	0	320
KwaZulu-Natal	Metro	Lower Income	13760	0	2560	320
KwaZulu-Natal	Metro	Middle Income	5440	2240	6080	1920
KwaZulu-Natal	Metro	Upper Income	960	0	960	1920
KwaZulu-Natal	Non-Metro Urban	Lower Income	8640	0	960	640
KwaZulu-Natal	Non-Metro Urban	Middle Income	4160	960	2240	1920
KwaZulu-Natal	Non-Metro Urban	Upper Income	320	0	320	640
North West	Non-Metro Urban	Lower Income	15040	640	640	1920
North West	Non-Metro Urban	Middle Income	5440	1920	960	3840
North West	Non-Metro Urban	Upper Income	640	0	0	1280
Gauteng	Metro	Lower Income	33280	2880	960	960
Gauteng	Metro	Middle Income	13440	3840	3520	5760
Gauteng	Metro	Upper Income	5120	640	1600	8000
Gauteng	Non-Metro Urban	Lower Income	4800	320	0	320
Gauteng	Non-Metro Urban	Middle Income	2240	320	320	1600
Gauteng	Non-Metro Urban	Upper Income	320	0	0	640
Mpumalanga	Non-Metro Urban	Lower Income	18240	320	960	1280
Mpumalanga	Non-Metro Urban	Middle Income	7680	960	960	3200

Mpumalanga	Non-Metro Urban	Upper Income	1600	0	320	2240
Limpopo	Non-Metro Urban	Lower Income	9920	0	320	640
Limpopo	Non-Metro Urban	Middle Income	5760	960	960	2240
Limpopo	Non-Metro Urban	Upper Income	1600	0	0	960

5. Weighting

Weights are provided in the data to account for the disproportionate sample allocation during data collection. The weights were generated using the StatsSA 2021 mid-year population estimates, for adults living in urban areas.

Weights were assigned to ensure the weighted sample records represent the target population as closely as possible. A weight indicates the number of population elements “represented” by a single sample element. Therefore, the sum of the weights should be equal to the relevant population total.

The weight of a respondent within a stratum is calculated as:

$$W_{pp} = \text{POP}_{\text{Stratum}} / n_{\text{Stratum}}$$

Where $\text{POP}_{\text{Stratum}}$ is the estimated population number of persons aged 18 and over in the stratum and n_{Stratum} is the total number of respondents in the stratum. The sum of the weights in the e-cigarette survey is 26 648 192, when we include all contacts, both contacted and non-contactable, which (other than rounding errors) corresponds to adults in urban South Africa (see Table 2).

6. Questionnaire

The survey instrument was developed through multiple rounds of review by the project team, Ask Afrika, and external reviewers. Ask Afrika used a market research application called Nebu to script the survey instrument. The efficacy of the scripted survey design and routing was tested internally by Ask Afrika and the TCDI project team.

The questionnaire had four main routing options, for the four respondents types:

1. Long-form routing for e-cigarette users (quota of 1250 respondents)
2. Long-form routing for cigarette-only users (quota of 700 respondents)
3. Short-form routing for cigarette-only users (applied once the quota of 700 respondents—above—had been filled)
4. Short-form routing for users of neither e-cigarettes nor traditional cigarettes

The first two options provide us with in-depth information about the respondents, whereas the second two options provide sufficient information for data users to conduct prevalence calculations.

There were two changes to the questionnaire that occurred after data collection had commenced. The details of these changes are outlined in section 9.3 below. The questionnaire can be found [here](#), and the differences in the questionnaire (before and after the questionnaire update) are outlined Table 8.

15% of interviews were subjected to quality assurance by Ask Afrika (see details regarding quality control in section 10).

7. Data collection plan and timeline

7.1 Initial timeline

The initial data collection plan was to collect the data in four phases, as well as a pilot phase. The plan was to conduct roughly an equal number of interviews per phase, equating to roughly 312 e-cigarette users per phase. The quota of 700 cigarette smokers was split across the phases, at 175 cigarette smokers per phase, to avoid collecting all 700 cigarette smokers early in the data collection process (since cigarette prevalence is far higher than e-cigarette prevalence).

7.2 Piloting

The first pilot survey was conducted between 8 and 15 December 2021. Interview recordings were reviewed and two key concerns were noted:

- Enumerators' understanding, pronunciation, and recording of a critical routing question (question 5) was problematic. This question determined whether a respondent was an e-cigarette user or cigarette smoker, both, or neither. Because of this, enumerators were debriefed on pronunciation and the nuances of the various questions, particularly question 5.
- Some data were missing due to incorrect routing by the CATI survey instrument. The script routing was adjusted to ensure that the survey instrument was operating as intended.

The data from the initial pilot phase were deemed unusable and discarded. The challenges with the initial pilot attempt meant that significant retraining of fieldworkers and adjustments to the survey instrument were undertaken. A second pilot round was therefore conducted between 14 and 17 January 2022. After a review of the data from this second round, it was agreed that the enumerators' performance had improved when asking the critical questions which determined the respondent's e-cigarette/smoking status.

However, the format of question 5 was still highly complex (resulting in enumerators often having to prompt respondents a few times to obtain clear answers). For this reason, the question was changed from:

Q5: "Have you ever smoked a standard cigarette and/or vaped, (i.e., used an electronic cigarette, heated tobacco product, or other vaping device) even if it was one puff?"

To two separate questions:

Q5a: “Have you ever used a vaping device (i.e., an e-cigarette, heated tobacco product, or other vape) even if it was one puff?”

And:

Q5b: “Have you ever smoked a standard cigarette even if it was one puff?”

Although the enumerators had to spend more time to obtain credible answers from respondents in this second pilot phase, the data that were collected were deemed credible, and for this reason these data were included in the first phase of data collection. For this data, users should note that respondents were asked the original question 5 (in one part), while all the rest of the data were asked question 5A and 5B separately. The pilot quotas were set at n=7 for cigarette-only respondents, and n=11 for e-cigarette respondents.

7.3 Data collection timeline for Phases 1 to 3

7.3.1. Phase 1

Phase 1 data collection was undertaken in the week of 17 to 24 February 2022. In the first three days of data collection, cigarette-only smoker surveys were completed at a far higher rate than those for e-cigarette users. This was due to the higher prevalence of cigarette smoking than e-cigarette use. Once the cigarette-only quota was filled, the scripting software only allowed the completion of e-cigarette surveys (the cigarette-only quota was closed).

After the completion of phase 1, the data were reviewed in detail. At this stage, there were some changes to the questionnaire. These are outlined in detail in section 9.2.

7.3.2. Phase 2

Phase 2 data collection was conducted between 31 March and 13 May 2022. As in phase 1, the cigarette-smoker quota was closed within the first weeks. The second phase moved slowly once the cigarette-only quota closed, with difficulties caused by South Africa’s daily occurrence of blackouts (loadshedding) at the time, as well as discouragement amongst fieldworkers, given the very low prevalence of e-cigarette users. This phase was ended on 13 May 2022, having not reached the desired e-cigarette user interview target.

7.3.3. Phase 3

Phase 3 started on 21 June 2022, and at this point it was agreed that the last two phases would be combined into one final phase, in the interest of time.

Data collector morale, although continuously monitored, was further impacted by loadshedding, which increased at this point and affected the fieldworkers’ ability to contact respondents. The impact of loadshedding was felt for most of the phase 3 fieldwork period, with varying degrees of severity, from 21 June until 7 August when loadshedding was suspended for a time.

The project team met weekly throughout the final phase to track fieldwork completion. The final enumeration was undertaken on 1 September 2022.

7.4 Sample management between phases

At the outset of the project, we had planned to conduct the data collection in four equal phases, and had anticipated a contact rate of 1:20 (contacts to telephone numbers drawn). As is described in section 4, we needed to make roughly 40 000 successful calls in order to reach the quota of 1250 e-cigarette users; with a contact rate of 1:20, this would require us to draw 800 000 contact numbers. Therefore, for phase 1, the sample allocation was divided into equal batches of 25%, and overlaid with the Credit Bureau sample, which provided the fieldwork call centre with 200 000 contacts numbers.

After phase 1 of data collection, there were sufficient contact numbers remaining (of the 200 000 contact numbers drawn in batch one) to proceed with the next phase. At this point, the data-processing team cleaned the first batch of any contacts that had reached a point of exhaustion (e.g., completed surveys, refusals, invalid numbers, or sample records that were over-dialled with 6 or more tries) and proceeded with the fieldwork using the remaining contacts from batch one, before drawing the next batch of contact numbers.

At the start of phase 3, the data-processing team again cleaned the initial batch of numbers of any contacts that had reached a point of exhaustion. At this point, Ask Afrika noted that a 1:10 contact ratio was sufficient to reach and survey the target of 1250 e-cigarette users, and a second batch of contacts (200 000 contacts) was procured using the sample approach as in phase 1, and these new contacts were added to the remaining, cleaned, sample. The re-using of the first batch of cleaned contacts—while simultaneously procuring the second batch of 200 000 contacts—allowed fieldworkers to resume phase 3 without the two-week delay that was usually required to procure a new batch of contacts.

Table 6: The three data collection phases, their sample allocation and project dates

Phase	Sample allocation	Project dates
Phase 1	25%	17-24 February 2022
Phase 2	25%	31 Mar-13 May 2022
Phase 3	50%	21 Jun-1 Sept 2022

8. Final sample

The survey team set out to collect 1250 completed interviews of people who had used e-cigarettes regularly (irrespective of the respondent's history with use of traditional cigarettes), and 700 completed interviews of cigarette-only smokers (respondents who were not e-cigarette users—see Table 1). The aim was to conduct a total of 1950 complete (long-form) surveys. While collecting these quota interviews, we also conducted short interviews of any cigarette-only smokers (over and above the quota of 700) and non-users of both products, for use in prevalence calculations.

Table 7 below outlines the final sample collected. Ultimately, Ask Afrika interviewed 1995 respondents, of which 1255 were e-cigarette users and 740 were cigarette smokers (only). Additionally, 19 268 short surveys were conducted in order to obtain prevalence estimates (including 16 658 non-users of e-cigarettes or combustible cigarettes and 2 610 smokers of combustible cigarettes).

A total of 399 810 contact numbers were drawn to reach these respondents. 83.1% of this overall sample were unreachable (i.e. numerous call attempts were made, but the respondents could not be reached), while 6.6% of the contact numbers were unusable (e.g., the number was blacklisted or out of order). 10% of the numbers tried were reached; of these, 4.8% resulted in complete quota surveys (n=1995), 10.2% were refusals (n=4248), 38.4% were incomplete surveys (n=15934), 40.2% were short surveys by respondents who were users of neither e-cigarettes nor combustible cigarettes (n=16658), and 6.3% were short interviews of cigarette-only smokers contacted after the cigarette-only quota had been filled (n=2610). The 10% contact rate (respondents reached) reflects the 1:10 strike rate discussed in section 6.4.

Table 7: Summary of data collected, by call outcome and user type

Call outcome	User type					Total	Share of all data
	Never user	Cigarette smoker	E-cigarette user	Dual user	Unknown/NA		
Reached							10%
Complete	0	740	367	888	0	1995	0,5%
Refused	0	0	0	0	4248	4248	1,1%
Incomplete	0	1256	47	80	14551	15934	4,0%
Non-user	16658	0	0	0	0	16658	4,2%
Smoker non-quota	0	2597	0	13	0	2610	0,7%
Not reached							90%
Unusable	0	0	0	0	26247	26247	6,6%
Unreachable	0	0	0	0	332118	332118	83,1%
Total	16658	4593	414	981	377164	399810	100%

9. Caveats and problematic variables

9.1 Initial routing question

In section 6.2, we described the change to a routing question during the pilot phase. The routing question Q5 (“Have you ever smoked a standard cigarette and/or vaped, (i.e., used an electronic cigarette, heated tobacco product, or other vaping device) even if it was one puff?”) was split into two separate questions, Q5a and Q5b.

This change from having a question (Q5) to two questions (Q5a and Q5b) had some implications for the questionnaire and data. For the second pilot survey (the responses to which were included in the phase 1 data), only Q5 was included; however, in phases 1-3, respondents were asked Q5a and Q5b, and then Q5 (which remained in the survey instrument to facilitate routing) was autocompleted using the Q5a and Q5b responses. For this reason, all three parts of the question are in the questionnaire, but respondents only provided responses to Q5a and Q5b, while Q5 was autocompleted by the software (respondents did not engage with Q5, except for the small number of respondents from the second pilot phase—see section 7.2). In the dataset, there are two separate variables for these responses, namely “usage_ever_ends” for Q5a and “usage_ever_cig” for Q5b.

9.2 Changes to the questionnaire

9.2.1. Refusals and non-users

After phase 1 of data collection, the data were reviewed and refusals appeared to be overstated, while non-users appeared to be understated. After listening to multiple surveys, the quality assurance team determined that in a number of refusal instances, the respondent had opted to “refuse” when in fact they were non-smokers/non-e-cigarette users. It was not possible to distinguish non-smokers from genuine refusals. Under-reporting non-users would result in over-estimated prevalence rates being calculated using the data.

For this reason, it was decided that the introduction in the survey questionnaire should include an additional answer option, “Never smoked/vaped”. The respondents who selected this option would then be asked to provide basic demographic details (age, sex, population group and the province in which they live).

Data users should note that phase one has an overstated refusal rate and an understated “non-user” rate. This is not the case for the data from the second and third phases of data collection.

9.2.2 Collection of data from cigarette smokers (quota vs non-quota)

The data collection plan included a quota (target) of 700 completed interviews of respondents with a history of regular cigarette use (but no history of regular e-cigarette use), referred to as “cigarette-only users. The quota of 700 cigarette-only users was staggered so that they were enumerated equally over the various phases of the survey. A detailed questionnaire was used to enumerate this group (see Table 8 below (code 2)).

In phase 1, after the quota of cigarette-smokers to interview was filled, cigarette-only users were identified by their answers to Q5a and Q5b and were then asked basic demographic questions, but no further questions relating to their smoking behaviour. At this point the interview was terminated (they were not asked module X; see code 3 in Table 8). However, this was problematic for cigarette prevalence calculations, as Q5b is insufficient to determine whether the respondent was a current regular smoker (they may have been a past regular smoker, or an experimental smoker).

Because of this, the routing of the questionnaire was updated for phases 2 and 3 of the survey: once the quota of cigarette-only users was filled, respondents who answered “No” to Q5a and “Yes” to Q5b were asked a few questions to identify their cigarette smoking status, so that they could accurately be included in cigarette smoking prevalence calculations. See code 4 in Table 8.

Table 8: Questions for respondent types, by quota and phase

Code	User type	Quota	Phase	Questions
1	E-cigarette user	Quota	1-3	All modules except X
2	Cigarettes only	Quota	1-3	All demographics QX1-QX7 Q32-Q39
3	Cigarettes only	Non/post-quota	1	Basic demographics (S1 - Q4) Q5a + Q5b
4	Cigarettes only	Non/post-quota	2+3	Basic demographics (S1 - Q4) Q5a + Q5b QX1-QX7
5	Never user	NA	1	Question 1v1 Basic demographics (S1 - Q4)
6	Never user	NA	2+3	Question 1v2 Basic demographics (S0 - Q4)

Because of this change in the questionnaire, the data records three different cigarette-user groups in the respondent outcome variable. The sample size for these three groups is shown in Table 9.

Table 9: Number of quota and post-quota smokers

Complete cigarette only surveys	Table 8 code	Number of observations
<i>Included in quota</i>		
Cigarette only user (phases 1, 2 and 3)	2	740
<i>Post-quota</i>		
Phase 1; pre-questionnaire change	3	479
Phases 2 & 3; updated questionnaire	4	2118*
Total		3337

Note: There were 13 “cigarette-only users” post-quota respondents who also filled in the e-cigarette modules (dual users). For this reason, there are 13 fewer respondents in this table than in Table 7, which has all 2610 post-quota cigarette smokers.

Users should use caution when analysing the phase 1, post-quota, data (N=479) for the purpose of prevalence calculations, as the data does not provide sufficient information to allocate these respondents to either the cigarette-smoker or the non-user group.

9.3 Date variable

The date variable (“lastdate”) included in the data reflects the date of the last attempt made by the fieldwork company to reach that number. For example, phone number X was drawn as part of the phase 1 sample batch, and there were four unsuccessful attempts made to contact number X during phase 1. During phase 2, there were a further two unsuccessful attempts to contact number X. The respondent was then coded as “Unreachable”, and “lastdate” reflects the date of the final unsuccessful call attempt. If an interview was interrupted and the respondent was called back at a later stage to complete the interview, the “lastdate” variable will reflect the later interview date.

The phase variable was generated and allocated to respondents after the completion of each phase. However, the number of observations will increase in later phases because sample units who were tried multiple times tended to move into later phases (particularly phase 3, even if the sample unit was originally drawn in an earlier phase). It is also worth noting that the “lastdate” variable was impacted by the quality assurance team’s interactions with the data (reviews); in these cases, the “lastdate” variable reflects the date of review. For observations that do have responses (“reached” respondents, as per Table 7), the “lastdate” variable is an approximation of the date of the last interview with the respondent.

The survey data were collected between 17 February and 1 September 2022 (with the exception of a small number of observations collected during the second pilot). However, considering the above caveats, analysis by specific date within the year, or by phase, is not advisable.

10. Quality control

All calls were recorded for purposes of quality assurance, with respondent permission. The recordings were saved on secure cloud storage. The in-field quality team was headed by a quality-assurance coordinator who selected surveys that the quality assessors were required to evaluate daily. During the in-field quality assurance process, 15% of each enumerator's completed interviews were checked. Checks included listening to the recordings and reviewing the data captured against the respondents' answers. Daily checks ensured that discrepancies could be identified and corrected without a significant impact on timing.

Quality-assurance reviews were conducted to check for:

- Incorrect respondent selection through deviations from the script,
- Non-persuasive and poor tonality of surveys,
- Data capturing errors.

Apart from the first pilot-phase data which were discarded (as mentioned in section 6.2), one interviewer was removed from the data collection operations because of not adhering to the script, and seven of his surveys were discarded.

11. Ethics clearance

Ethics clearance for data collection was granted by the University of Cape Town Faculty of Commerce Faculty Ethics in Research Committee. The ethics approval reference is REC 2021/10/001.