CVACS

COVID-19 VACCINE SURVEY

Panel User Manual

Survey 2

Release April 2022

Timothy Brophy
Kim Ingle
Brendan Maughan-Brown
Katherine Eyal
Alison Buttenheim

SALDRU Southern Africa Labour and Development Research Unit
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Contents

1 What is the COVID-19 Vaccine Survey?

The COVID-19 Vaccine Survey (CVACS) is a South African national panel study of individuals initially unvaccinated against COVID-19. CVACS is implemented by the Southern Africa Labour and Development Research Unit (SALDRU) based at the University of Cape Town's (UCT's) School of Economics.

The same respondents were interviewed twice, a few months apart, to gather information about their attitudes, beliefs and intentions regarding COVID-19 vaccination. The purpose of CVACS was to collect high quality, timely, and relevant information on facilitators and barriers to COVID-19 vaccine uptake - including vaccine hesitancy and access constraints - to contribute to the development of data-driven campaigns and programmes to increase COVID-19 vaccination uptake in South Africa.

Multiple outcomes are measured by CVACS, such as vaccination knowledge, attitudes, and beliefs (including specific concerns); vaccination intentions (plans to get vaccinated), vaccine access constraints, and information seeking.

To ensure that the most policy-relevant information was collected, the questionnaires were co-created with academics, policymakers, and members of civil society organizations involved in the COVID-19 response and vaccination roll-out. Co-development of questions was achieved through workshops, one-on-one engagements and policy crowdsourcing.

Survey 2 was conducted in February and March 2022. Survey 2 attempted to reinterview the same Survey 1 respondents and asked questions of both individuals who remained unvaccinated and those who had been vaccinated since Survey 1. Further unvaccinated respondents were recruited to replace Survey 1 respondents who dropped out of the survey, in order to achieve a sample of unvaccinated respondents of similar size to Survey 1.

CVACS was not designed to be, and should not be used as a prevalence study. The data cannot be considered to be nationally representative of all unvaccinated individuals in South Africa.

2 This manual and how to cite it

This manual is designed to assist users of the CVACS dataset. It contains information about the design, operations and data for both Survey 1 and Survey 2.

This manual was prepared using Version 1.1.0 of the CVACS Survey 1 dataset and Version 1.0.0 of the CVACS Survey 2 dataset.

Users wishing to cite this manual should use the following citation:

Brophy, T., Ingle, K., Maughan-Brown, B., Eyal, K., Buttenheim, A. COVID-19 Vaccine Survey (CVACS) 2022: Panel User Manual. Cape Town: Southern Africa Labour and Development Research Unit.

3 Correct citation of the CVACS data

Users wishing to cite the latest versions of the Survey 1 and Survey 2 datasets should use the following citations:

Survey 1:
COVID-19 Vaccine Survey (CVACS) 2021, Survey 1 [dataset]. Version 1.1.0. Cape Town: Bill and Melinda Gates Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2021. Cape Town: DataFirst [distributor], 2022. DOI: https://doi.org/10.25828/82ww-5r26

Survey 2:
COVID-19 Vaccine Survey (CVACS) 2022, Survey 2 [dataset]. Version 1.0.0. Cape Town: Bill and Melinda Gates Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2022. Cape Town: DataFirst [distributor], 2022. DOI: https://doi.org/10.25828/Y6FM-1K12

4 Sampling and weights

4.1 Sampling

4.1.1 Survey 1 sampling

The CVACS Survey 1 sample was obtained in two steps. First, a stratified sample was drawn from the GeoTerraImage (GTI) 2021 sampling frame (https://geoterraimage.com/), using individuals aged eighteen and older. The sample was primarily stratified across the following categories: province, population group, geographic area type (metro, non-metro urban, non-metro rural) and the neighbourhood lifestyle index (NLI), in groups of NLI 1-2, NLI 3-4, and NLI 5-10. The NLI is an income-based segmented classification of neighbourhoods according to their income and various lifestyle characteristics at various spatial levels. Age categories defined according to the COVID-19 vaccination roll-out age groups (18-34, 35-49, 50-59, 60+), and gender were used as further explicit stratification variables to ensure sufficient coverage across the categories of interest. In total 508 strata were formed. The sample was disproportionately allocated to the strata.

Second, a Credit Bureau database (containing more than 16 million cell phone numbers) was then linked to the sampling frame at the Enumeration Area (EA) level. The Credit Bureau database contains contact details of individuals who have had a credit check. The reason for the credit check could be for a credit application, rental application, job application or other use. The outcome of the credit check had no consequence on the individual's eligibility to be included in the sampling frame. Within each stratum, a random sample of cell phone numbers (respondents) was selected for the computer-assisted telephone Interview (CATI) survey.

The CVACS sample aimed to be as representative as possible. Several limitations and potential sources of bias should be considered when interpreting the findings:

- Individuals who have not had a credit check do not appear in the credit bureau database. The credit bureau database is thus not representative of all South Africans.
- CVACS is a telephone survey - access to a telephone is an eligibility criterion. Individuals with access to a telephone, primarily a cellphone, likely have different characteristics to the overall South African population, including greater access to resources.

- Non-response for telephone surveys is typically higher compared to in field surveys. Adjustments for non-response are made within each of the 508 strata. However, these adjustments assume that the individuals who respond within a stratum are not systematically different to the non-respondents. With very high non-response rates, the weighting adjustments are unlikely to account for all select bias in non-response.
- Self-report of not having been vaccinated against COVID-19 was an eligibility criterion for the survey. That is, CVACS Survey 1 interviewed only those individuals who said they were unvaccinated against COVID-19. A total number of 14,577 individuals were screened on vaccination status: 10,536 individuals reported being vaccinated, 3,969 reported being unvaccinated, 64 refused and 8 reported "don't know." The proportion who reported that they were vaccinated (72%) is significantly greater than the proportion of the adult population of South Africa vaccinated at the time of the CVACS Survey 1. This is suggestive of significant bias in self-reporting of vaccination status.

Given the high proportion of individuals reporting that they were vaccinated, we adjusted the recruitment strategy at the start of the fieldwork period in order to minimise self-reporting bias of vaccination status. Instead of confirming the respondent's name at the start of the interview, respondents were invited to share his/her name at the end of the survey. Consequently, respondents reported their vaccination status before sharing their name. While the vast majority of individuals interviewed were the exact individuals selected from the sample frame, some individuals who did not meet the initial sampling criteria were included in the survey.

The final realized sample was approximately balanced by gender, and had a similar provincial distribution to the South African population. The observed distribution across the vaccination roll-out age categories was slightly more skewed to those in younger age categories than originally intended, reflecting the higher rates of vaccination among the older South African population.

4.1.2 Survey 2 sampling and top-up sample

Survey 2 attempted to re-interview the 3,510 successfully interviewed Survey 1 respondents. A total of 1,772 Survey 1 respondents were successfully re-interviewed in Survey 2. Among these, 386 self-reported being vaccinated between Survey 1 (November & December 2021) and Survey 2 (February & March 2022), with a further 1,386 self-reporting that they remained unvaccinated in Survey 2.

In each survey round, CVACS aimed to realise a sample size of approximately 3,500 unvaccinated respondents. In Survey 2, we knew that some respondents would become vaccinated between Survey 1 and Survey 2 and, as a consequence, they would be excluded from the final unvaccinated sample. In addition, between Survey 1 and Survey 2, we anticipated a loss of respondents due to attrition. To reach the target sample of unvaccinated respondents of approximately 3,500, we replenished the sample via a sample top-up by randomly drawing new respondents from the same sampling frame as Survey 1. These new respondents were selected from the same strata as those Survey 1 respondents that had become vaccinated or attrited between survey rounds[1]. See Survey 1 sampling for more detail on the Survey 1 sampling frame.

[1] There were a few respondents who were not selected from the exact same strata as someone who had dropped out due to timing of sample replacement versus determining final outcomes of sample respondents during fieldwork for example.

A top-up sample totalling 14,067 individuals were selected from the sample frame for attempted contact by the fieldworkers for Survey 2. Of these 14,067 individuals, we were able to contact 2,222 individuals who told us that they were unvaccinated and who agreed to participate, resulting in the following total sample for Survey 2:

Table 1: Composition of CVACS Survey 2 sample

Sample		Frequency	Percent
Survey 1 CVACS sample	Unvaccinated1	1,386	34.7%
	Vaccinated	386	9.7%
Survey 2 top-up sample	Unvaccinated1	2,222	55.6%
Total		3,994	100.0%

[1]A total of 3,608 unvaccinated respondent interviews were realised in Survey 2.

Contained in the Survey 2 datasets is a variable *s2_{i}_sample* (where {i} denotes the data file). This variable shows whether the sample member is from the original CVACS Survey 1 sample or the CVACS Survey 2 top-up sample.

The final interview outcomes for Survey 1 and Survey 2 are as follows:

Table 2: Survey 1 and Survey 2 interview outcomes

Outcome	Survey 1		Survey 2	
	Frequency	Percent	Frequency	Percent
Successfully interviewed	3,510	100	3,994	69.68
Refused/ not available			178	3.11
Not contacted in Survey 2			1,510	26.34
Not tracked in Survey 2			45	0.79
Moved outside of South Africa			3	0.05
Deceased this wave			2	0.03
Total	3,510	100	5,732	100

4.1.3 Sample participation and non-participation over the waves

Participation in surveys is voluntary, so there is likely to be attrition between longitudinal survey rounds. There are many reasons for respondents not to participate in a follow-up round, including refusals to participate, inability to contact respondents, respondents moving outside of South Africa, and death of respondents. In a recent study, NIDS-CRAM, a South African telephonic survey based on the National Income Dynamics Study(NIDS[2]), had an attrition rate of 20% between its first two rounds of telephonic data collection in 2020 (Daniels et al., 2022), whereas CVACS attrition was approximately 49% between Survey 1 and Survey 2 Although high, this is perhaps less surprising when considering the sensitive nature of the questions asked in CVACS. According to Daniels et al. (2022) the quality of phone numbers, the number of phone numbers per respondent, and alternate contact numbers matter greatly in reducing attrition rates. As part of Survey 1, CVACS collected multiple phone numbers per respondent, but this was contingent upon respondents' willingness to provide multiple contact numbers. Unlike NIDS-CRAM, CVACS also did not have the benefit of being based on a sample of an established

[2]NIDS is a long running longitudinal survey from 2008 with 5 waves of data collection, and thus a very well established panel.

panel survey; it also did not have a database of multiple phone numbers at the start of the study for most respondents, to draw on as alternative contact numbers during the study.

The table below shows the Survey 1 respondents' participation in Survey 2.

Table 3: Sample participation and non-participation of Survey 1 participants in Survey 2

Survey 2 outcome	Frequency	Percent
Successfully interviewed (balanced panel)	1,772	50.48
Refused/ not available	178	5.07
Not contacted in Survey 2	1,510	43.02
Not tracked in Survey 2	45	1.28
Moved outside of South Africa	3	0.09
Deceased this wave	2	0.06
Total	3,510	100

Survey 2 of CVACS records a very low death rate since Survey 1. To reliably count deaths between survey rounds, a relative or close friend ofeach deceased respondent would have to answer the telephone and report the death. However, this is unlikely to happen in all cases. A deceased respondent will quite possibly be recorded as not contacted instead, especially when considering the high proportion of the Survey 1 sample who were not contacted. Because of this, CVACS death rates should not be used in any analysis.

4.2 Weighting

Weights were assigned to make weighted sample records as representative as possible as the sampling frame.

Weights were developed to compensate for

- Unequal inclusion probabilities due to the sample design, by calculation of the design weights.
- Non-response within each stratum, by adjusting the design weight, if necessary.

4.2.1 Survey 1 design weights – corrected for nonresponse

All unvaccinated respondents and those who indicated that they are vaccinated were considered for the weighting process. The design weight for a respondent in stratum *i*, is calculated by

$$w_{PP_i} = \frac{N_i}{n_i}$$

where N_i is the 18+ population number and n_i the number of respondents selected for the sample in stratum *i*.

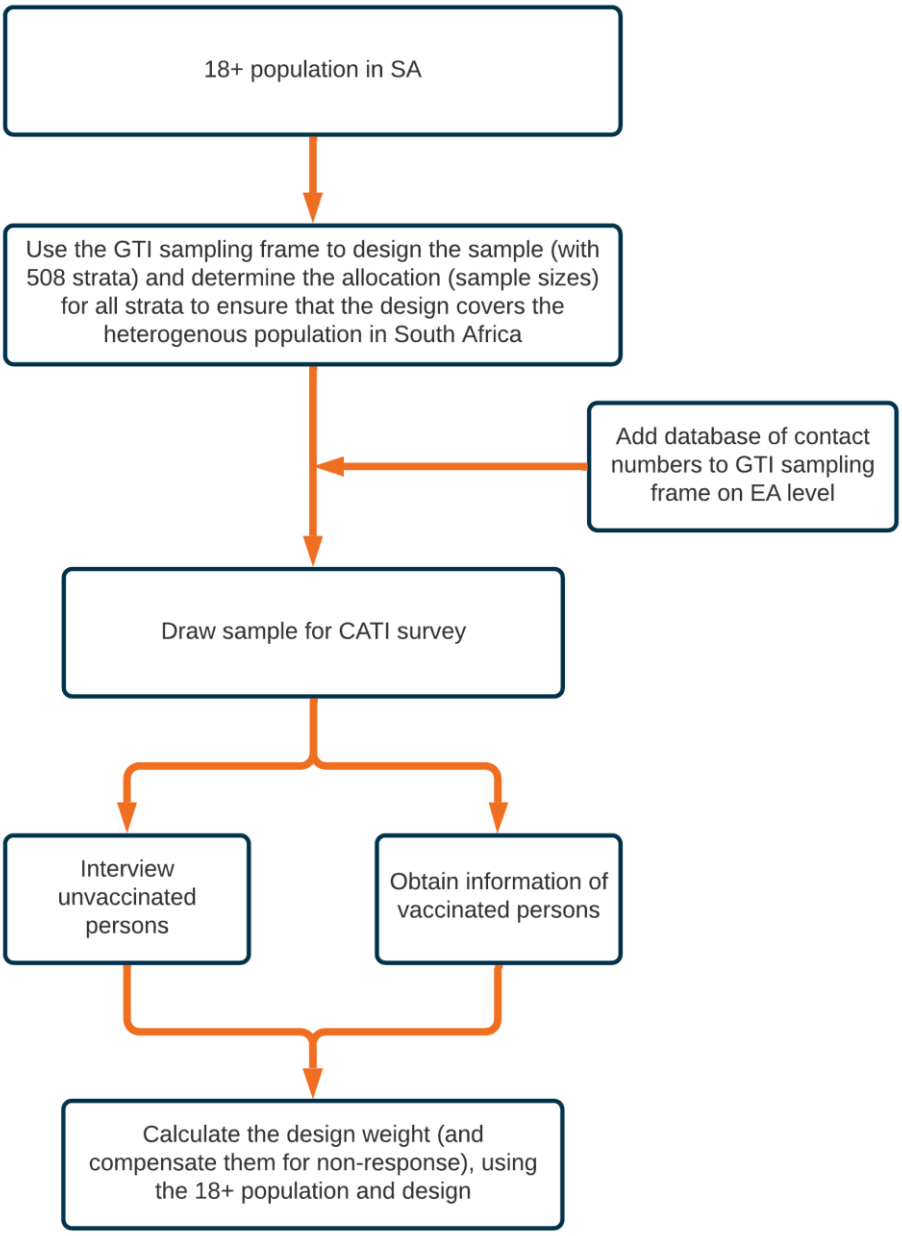
Hereafter, the design weights are adjusted to compensate for non-response, within each stratum, as follows:

$$w^*_{PP_i} = \tau_i w_{PP_i}$$

where $\tau_i = \frac{sum\ of\ weights\ of\ designed\ number\ of\ respondents\ in\ stratum\ i}{sum\ of\ weights\ of\ realised\ number\ of\ respondents\ in\ stratum\ i}$, the non-response weight factor.

Please see the diagram below for a description of the Survey 1 sampling and weighting process.

4.2.2 Illustration of the Survey 1 sampling and weighting process



4.2.3 Survey 2 weighting

The CVACS Survey 2 data includes design weights for the full Survey 2 sample of unvaccinated and vaccinated individuals. Panel weights are included for the sample of individuals from Survey 1 who were re-interviewed in Survey 2, whether vaccinated or unvaccinated. Weighted CVACS data, whether design or panel-weighted, is not nationally representative of all unvaccinated individuals in South Africa.

The Survey 2 design weight corrects for non-response and uneven inclusion probabilities within the sampling frame. The panel weight corrects for a degree of attrition bias in the CVACS balanced panel data. With high rates of non-response and attrition, the weighting adjustments are unlikely to account for all selection bias in non-response or self-reported vaccination status or all attrition bias across Survey 1 and Survey 2.

4.2.4 Survey 2 design weights

Carry-over weights were used for the Survey 1 panel sample who were re-interviewed in Survey 2 - their Survey 1 design weight is used as their Survey 2 design weight (i.e., their design weight is the same in Survey 1 and Survey 2. The sum of weights of the panel members in Survey 2 stratum *i* is defined by N_i^p . The remainder of the population total in the stratum is equal to

$$N_i^R = N_i - N_i^p$$

where N_i is the 18+ population number in stratum *i*, according to the GeoTerraImage (GTI) 2021 population estimates.

The weights for the top-up sample respondents – vaccinated or unvaccinated, were calculated similarly to the design weights in Survey 1:

The design weight for a top-up sample respondent in stratum *i*, is calculated by

$$w_{TU} = \frac{N_i^R}{n_i^R}$$

where n_i^R is the number of top-up sample respondents selected in stratum *i*.

Hereafter, the design weights are adjusted to compensate for non-response, within each stratum, as follows:

$$w_{PP_i}^* = \tau_i w_{TU_i}$$

where $\tau_i = \frac{sum\ of\ weights\ of\ the\ designed\ top{-}up\ sample\ respondents\ in\ stratum\ i}{sum\ of\ weights\ of\ realised\ top{-}up\ sample\ respondents\ in\ stratum\ i}$,
the non-response weight factor. $w_{PP_i}^*$ for Survey 2 panel respondents are the (carry-through) Survey 1 design weight.

4.2.5 Panel weights: Survey 1 to Survey 2

Panel surveys invariably suffer from attrition and CVACS is no exception. Table 3 shows the Survey 2 outcomes for the 3,510 individuals who were successfully interviewed in Survey 1. The attrition rate between the two surveys was around 49%. Individuals who attrite tend to be systematically different to

individuals who remain in the panel and panel weights are necessary to correct for bias due to this non-random attrition.

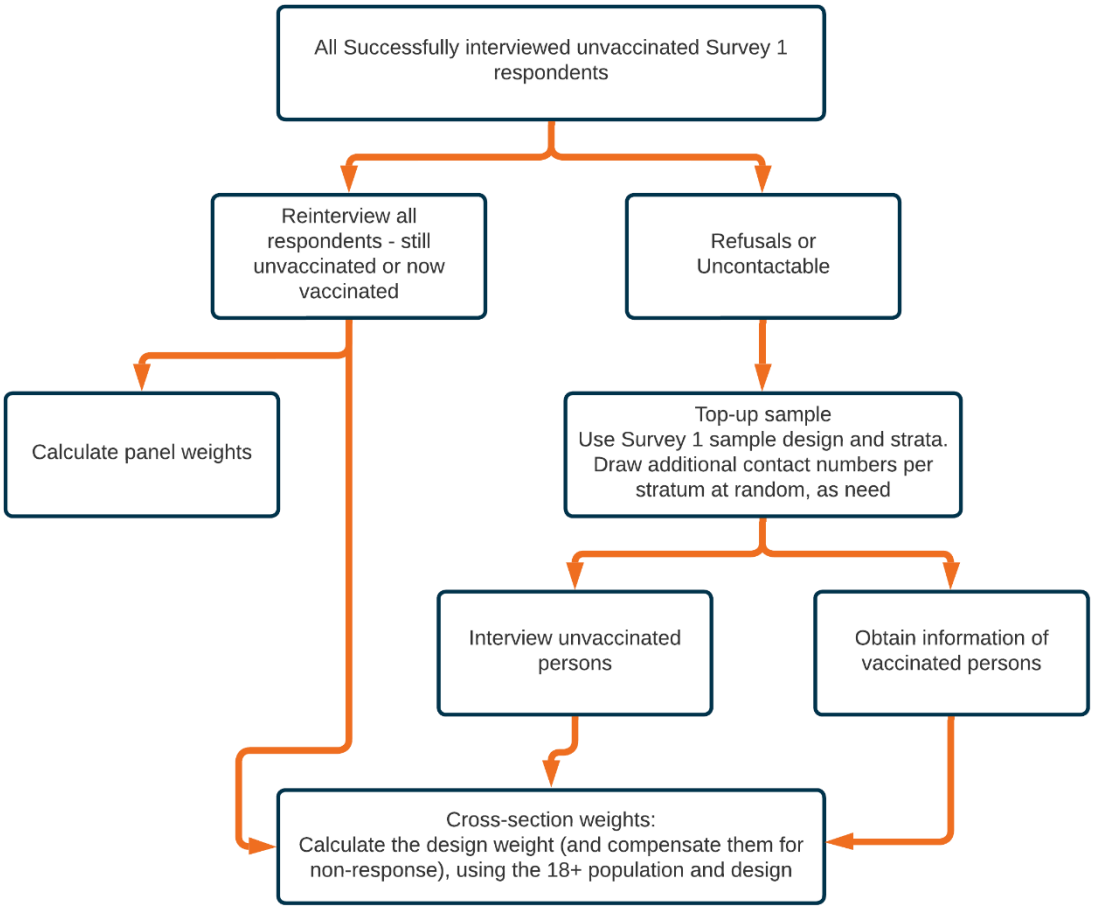
The first step in creating the panel weights for Survey 2 is to model the probability of being re-interviewed in Survey 2, which is denoted as $\hat{p}_i$. The following explanatory variables were included in the final probit model – province, population group, geographic area type (metro/non-metro urban/non-metro rural), income level (using the NLI categories), indicators for the vaccine roll-out age categories, gender and race[3]. The design weights of the unvaccinated respondents in Survey 1 were used as the weight variable in the probit analysis, to account for non-response or uneven inclusion probabilities in Survey 1. In this multivariable model, attrition was statistically significantly higher among women, higher-income individuals, and younger individuals.

The panel weight for Survey 2 is the product of the Survey 2 design weights and the inverse of the conditional probability of being re-interviewed, calculated as follows:

$$Panel\ weight\ =\ w^{*}_{PP_i} * \ 1/\ \hat{p}_i$$

Please see the diagram below for a description of the Survey 2 sampling and weighting process.

4.2.6 Illustration of the Survey 2 sampling and weighting process



[3] These are the components of the design weights, and were found to be the strongest predictors of attrition.

4.2.7 Using the CVACS weights in Stata

In Stata the command to set the survey design in CVACS Survey {x}
(where {x} is the number of the survey round) using the design weight is:

```
svyset [weight = s{x}_dwgt]
```

In Stata the command to set the survey design in the balanced panel using the panel weight is:

```
svyset [weight = s2_bp_pweight]
```

5 Data collection

5.1 Method of data collection

A single survey instrument was administered for CVACS Survey 1 data collection using computer-assisted telephone interviewing (CATI).

The realized Survey 1 sample consisted of 3510 individual interviews, all of whom self-reported that they were unvaccinated for COVID-19.

With Survey 2, we administered two instruments via CATI. The first instrument was administered to all unvaccinated respondents - those panel respondents that remained unvaccinated between Surveys 1 & 2 and the sample top-up members. The second instrument was administered to those respondents that reported being vaccinated between Surveys 1 & 2.

Survey 2 realised a sample of 3,994 respondents consisting of 3608 unvaccinated and 386 vaccinated respondents.

The CVACS questionnaires for both Survey 1 and Survey 2 were translated into all official South African languages and interviews were conducted in the preferred language of the respondent. In both survey rounds, most of the survey questions collected individual-level data, with some household-level data also collected through the individual questionnaire.

5.2 Informed consent and ethics approval

CVACS was approved by the Ethics in Research Committee (EiRC) of the Faculty of Commerce, University of Cape Town (See appendix Appendix – CVACS ethics approval). Informed consent was obtained from all respondents. Respondents were informed of their rights to refuse to answer the questionnaire entirely or to refuse to answer any particular question; and that their data will remain confidential. The Survey 1 informed consent statement, for example, was:

Informed Consent
Thank you for taking our call. We are doing important research about the wide variety of South African adults' opinions about COVID-19 and vaccinations. We want to learn about people's opinions and choices and how they might change over time. There are no right or wrong answers. We respect all preferences and choices. Your feedback will remain confidential and will never be linked to your name. Participation is voluntary and you can stop at any time. If you agree we might contact you in a few months' time to follow up. To thank you for your time you will be entered into a lucky prize draw. This call will be recorded for quality control. The survey will take approximately 10 minutes and is called the COVID-19 Vaccine Survey. UCT Commerce Ethics Committee approved the research study and you can reach us on 021-650-xxxx with any questions.

5.3 Data collection dates

Table 1 contains the start and end dates of the data collection periods for Survey 1 and Survey 2.

Table 4: Data collection dates

Survey	Start date	End date
1	15 November 2021	15 December 2021
2	23 February 2022	25 March 2022

5.4 Contact attempts

5.4.1 Call attempts

Phone numbers for the panel respondents were provided to the fieldwork company at the start of Survey 2. These phone numbers could belong to either the respondent or someone who might know how to contact the respondent.

The fieldwork company attempted to contact respondents using all phone numbers they were provided with at the beginning of the survey round, as well as any new numbers that they were provided by respondent contacts during the survey round. Each phone number was attempted multiple times before the fieldwork company stopped attempting to make contact with the respondent and mark them as "Not contacted in Survey 2".

5.4.2 SMSs

Before Survey 1 respondents were called for Survey 2, they were sent an SMS which thanked them for their participation and informed them that they would be called for the second round of interviews in the coming days.

When operational feasibility allowed for additional call attempts, sample members who were proving difficult to get a hold of (and could qualify or were about to qualify as “Not contacted in Survey 2”), were sent an additional SMS saying that we had been trying to reach them and that we would try to call them again, before contact was reattempted.

The number to send an SMS to was chosen by SALDRU. The number that SALDRU believed to be the most likely to result in a successful interview, was selected for this purpose. These were selected from the list of numbers that SALDRU provided the fieldwork company with at the start of Survey 2.

5.5 Data quality

Data quality assurance addressed the following:

5.5.1 Unit non-response

Unit non-response was mitigated via several measures:

- **Call rules:** CVACS initiated a call rule which indicated that there needed to be multiple attempts to contact the respondent before we would consider the number to be uncontactable.

- **Lucky draw incentives:** To thank CVACS respondents for their time, respondents in each survey round were informed that they would be entered into two separate lucky prize draws. One lucky draw occurred at the end of each survey round with three winners, each receiving a R3,000 gift voucher from a store of their choice. Additionally, each time they participated in a CVACS survey, respondents were entered into a grand prize draw which occurred after both survey rounds in April 2022, where they could win a R 10,000 in-store voucher.

5.5.2 Item non-response

Item non-response can arise for different reasons, for example when a respondent refuses to answer a question or does not know the answer. In such instances the “Don’t Know” or “Refused” response options are coded accordingly (See section 6.6 Non-response codes), allowing data users to estimate item non-response rates for relevant questions.

The use of CATI reduces the instances of interviewer-induced item non-response because CATI automates the skip pattern for the interviewer and prompts them if a question is missed. There was a strict policy that data were only accepted from data collection if all sections of the questionnaire had been answered.

5.5.3 Data consistency

In addition to item and unit non-response, internal consistency of the data within the instruments and against the sampling frame were also conducted. Involving several checks and mitigations to ensure data consistency:

- **Within-survey consistency checks in the office:** SALDRU carried out a range of pattern searches and consistency checks on the data during fieldwork to identify interviewer effects and possible miscapture. When areas of concern were found, the audio recording of respondents' interviews was reviewed by the quality control team to ensure that the data is correct.

- **Immediate correction** of **interviewer-related errors during the fieldwork period:** CATI allows data quality to be checked from the start of data collection by each interviewer. SALDRU was able to inform the data collection company of any unusual interviewer behaviour for immediate correction and retraining. Performing this in a timely manner enabled CVACS to mitigate and correct most interviewer effects.

6 Data files and structure

CVACS collected individual-level, and some household-level data through the individual questionnaire. Survey 1 was administered to unvaccinated respondents only. Survey 2 was administered to vaccinated panel respondents, as well as unvaccinated respondents from both the panel and top-up samples.

6.1 Data format

Data files are provided in Stata 14 file format but can also be opened with more recent versions of Stata.

6.2 Data file name structure

Each respondent was assigned a unique personal identifier that is used for both Survey 1 and Survey 2. This unique personal identifier is contained in the variable called *pid*.

Each record contained in the data files is unique on *pid* and merge to each other uniquely in a one-to-one relationship

All data files contained in the CVACS dataset have the following naming convention:

[Questionnaire]_CVACS_[Survey]_Anon_V[x.x.x]

Questionnaire: Describes the data contained in the data file. This is derived from the name of the questionnaire administered.

e.g: *Unvaccinated*

Survey: Indicates which survey round the data file contains.

e.g: *S1* indicates Survey 1

Anon: Indicates that all the Personal Identifiable Information (PII) data has been removed from the data file.

e.g: *Anon*

Vx.x.x: Represents the dataset version number. CVACS versions at the dataset level and not at the data file level. See section 6.3 Versioning of the dataset.

e.g: V1.0.0

6.3 Versioning of the dataset

The versioning of the CVACS dataset aims to inform users of major, minor, and patch fix releases of the data to ensure that they use the most up-to-date data for their research.

Version[major releases].[minor releases].[patch releases]

Major release: Refers to the number of a major release. A major release only occurs when a full release of the entire dataset is necessary. This normally only coincides with the release of another round of data collected.

e.g., V1.0.0, here 1 indicates the first major release of the dataset.

Minor release: Refers to the number of a minor release. A minor release is the release of a single data file that occurs between major releases, i.e., a release of corrected data where no new data has been collected.

e.g., V1.2.0, here the number 2 indicates a second minor release between major releases.

Patch releases: Refers to the number of a patch release. Patches are scripts (Stata do files) that are created to correct specific errors in the data. The scripts are given to users to run on the existing data files to fix a specific data issue. Once run, the patch will generate a new minor release version of the data.

e.g., V1.2.3, here the number 3 indicates that 3 patch scripts were released to fix data issues in version 1.2 of the dataset.

6.4 Data files

The data files that make up the CVACS datasets are as follows:

Unvaccinated: This is the data file containing data from responses to the questions in the questionnaire for unvaccinated respondents.

One record per respondent (Unique identifier: *pid*).

This data file does not contain any variables which could comprise anonymity. As a result, there are a few questions that appear in the questionnaire for which there is no corresponding data in the public release version.

Data file included in Survey 1 and Survey 2.

Vaccinated: This is the data file containing data from responses to the questions in the questionnaire for vaccinated respondents.

One record per respondent (Unique identifier: *pid*).

This data file does not contain any variables which could comprise anonymity. As a result, there are a few questions that appear in the questionnaire for which there is no corresponding data in the public release version.

Data file included in Survey 2.

derived: This data file contains derived variables that are not collected through questions asked directly of the respondent but are calculated or imputed from their responses to other questions.

One record per respondent for completed questionnaires only (Unique identifier: *pid*).

Data file included in Survey 1 and Survey 2.

Link File: This data file contains interview outcomes for each respondent in CVACS Survey 1 and Survey 2. This data file also includes data on those who died at some stage during the panel. It is thus useful to reconcile the data for all the respondents across the panel. To identify individuals who were added in the CVACS Survey 2 top-up, the variable *sample* is included in the Link file. This variable identifies which sample individual respondents originated from.

One record per panel respondent (Unique identifier: *pid*).

Data file included in Survey 2.

Since Stata limits variable labels to a length of 80 characters, the variable labels are based on a shortened version of the question asked to respondents. It is recommended that data users refer to the questionnaire for the relevant data file to review the full wording of the questions.

6.5 Derived data file

Certain variables are created by the CVACS team. These variables appear in the derived data file.

Derived variables are:

- Any variable that is finalised after data collection through a post-coding exercise;
- Any variable that is the result of a combination of other variables;
- Any variable that is imputed.

The derived variables include:

- Weights
  - The weights are discussed in section 4.2 Weighting.
- “Best” variables
  - These are the best estimates of certain demographic characteristics of the respondent. These are derived by considering responses by a respondent across the CVACS panel, and at times across various questions as well.

- Age intervals
  - Age intervals are calculated from the respondent's best age.

- National lockdown level at the time of interview
  - During the COVID-19 pandemic, the South African government decided upon different levels of "lockdown" (restriction of activities), in response to COVID-19 infections at the time, with the aim of reducing the spread of COVID-19.

- Employment status
  - The respondents' employment status is derived from various questions in the labour market section of the questionnaire. A Stata do file accompanies the release of the data which shows how this employment status variable was derived.

- Derived geographic variables
  - These geographic variables are derived from the area in which the respondent reported they were living at the time of the interview.

- Interviewer information
  - This includes basic demographic, language and experience information on the enumerator who conducted the interview for each respondent.

6.6 Non-response codes

Non-response codes are usually indicated by negative numbers. These codes are detailed below.

Table 5: Typical non-response codes in CVACS

Type of Item Non-Response	Non-Response Code
Don't know	-9
Refused	-8
Not applicable	-5
Missing	-3

-3 "Missing" is a data quality code which is only used post-fieldwork. It is applied to show the data user that a question should have been answered during fieldwork but was not for that observation.
Researchers are reminded to take these non-response codes into account when conducting their analysis.

6.7 Variable naming convention

Variable names use a prefix to specify a) the survey round (Survey 1 or Survey 2), and b) the data file:

[Survey round]_[Data File]_

Survey round: *"S{i}_"* where {i} represent the survey round number.
E.g., "S1_" survey round 1

Data file: *"_{x}_"* where {x} represents the data file in which the variable can be found.
E.g., "S1_u_" where "u" represents the Unvaccinated data file. Or
"S2_v_" where "v" represents the Vaccinated data file.

6.8 Surveyed vs. historical data

In Survey 2, certain questions were not re-asked of some respondents. This was done to avoid re-asking respondents' time-invariant data that we had collected previously. Where this was the case, the previously collected historic data was added back to the Vaccinated and Unvaccinated data files. For example, we assumed that self-reported population race group would not have changed between surveys. In order for users to differentiate between this historical data and the data which was collected in Survey 2, flag variables have been created. An example of this is *s2_u_popgrp_flg.*

In Survey 2, we had originally included a question asking respondents if they had moved since we last spoke to them in Survey 1. If they indicated that they had not moved we did not ask the type of area that they live in *(s2_{x}_are)* or the area description *(s2_{x}_area_des)* as that would not have changed since they had not moved. However, this question's implementation failed early during data collection. Consequently, for 235 respondents, 195 unvaccinated and 40 vaccinated, the data for the area and area_des variables are based on data collected in Survey 1. For all other Survey 2 respondents, data for these variables was collected in Survey 2.

7 Merging the data within a survey round

In order to merge observations for individuals from different data files, users should merge on the person level identifier, which is the variable *'pid'*. The relevant Stata code for merging version 1.1.0. of the Survey 1 data is as follows:

```
use "Unvaccinated_CVACS_S1_Anon_V1.1.0.dta", clear
merge 1:1 pid using "derived_CVACS_S1_Anon_V1.1.0.dta"
drop _merge
```

The relevant Stata code for merging version 1.0.0. of the Survey 2 data is:

```
use "Unvaccinated_CVACS_S2_Anon_V1.0.0.dta", clear
append using "Vaccinated_CVACS_S2_Anon_V1.0.0.dta"
merge 1:1 pid using "derived_CVACS_S2_Anon_V1.0.0.dta"
drop _merge
```

8 Important data limitations

8.1 Individual and household-level analysis

The limits of telephonic surveys with respect to questionnaire length and complexity is a key factor for researchers to take into account when performing their analysis. The household-level data are reported by the respondent on behalf of the household. It is not known whether the respondent was the most knowledgeable person in the household in terms of answering household level questions. It is therefore recommended that researchers avoid using CVACS to conduct household-level analysis. However, it is possible to estimate statistics at an individual level about household living conditions. Similar to how Kerr et al., (2020) explain for NIDS-CRAM, it is not legitimate to estimate that Y% of households had an internet-enabled device. One could say that Z% of adults live in households where members went hungry, but not that A% of households had members going hungry or that B% of members went hungry.

8.2 Regional analysis

Due to the size of the sample, we caution against drawing general inferences and conclusions at sub-national levels - this includes provincial or lower levels of geography.

8.3 Derived and imputed variables

8.3.1 Geographical variables

All CVACS derived geographic information is based on the area that the respondents self-reported they live in.

8.3.2 Monetary values

When comparing monetary values over time, inflation should be taken into account. Since Statistics South Africa calculates inflation rates each month, the monetary values from different months can be inflated/deflated to take into account inflation.

To help researchers control for inflation, deflator Stata do files accompany the releases of the data.

8.3.3 Open-ended questions

The CVACS questionnaire includes several open-ended questions where the enumerator recorded their translation of the response. This has the potential to introduce error which is difficult to quantify. Researchers can consult the questionnaire to view the question as it was asked of the respondent, together with any question-specific instructions for the enumerator.

9 Program library

Two kinds of coding files are provided: (i) those that give insight into derived variables, and (ii) those that assist with data manipulation of the survey data files.

9.1 Derived variables

9.1.1 Employment status

Script filenames: *Program1a-CVACS_Survey1_Employment_Status.do* & *Program1b-CVACS_Survey2_Employment_Status.do*

Employment status is coded using the International Labour Organisation's definitions to assign respondents to one of the following categories: not employed (i.e. unemployed or not economically active) or employed.

The respondent is determined to be employed if they are economically active and reported having any form of employment in the past week at the time of the interview.

9.1.2 Deflator

Script filenames: *Program2a-CVACS_Survey1_Deflating_Do_File.do* & *Program2b-CVACS_Survey2_Deflating_Do_File.do*

Because data collection for CVACS took place over multiple months, all financial data need to be deflated.

9.2 Data manipulation – merging datasets

It should be noted that, in general, merges should always be done on the personal identifier (*pid*) which is the unique identifier across all data files for all rounds of the survey.

9.2.1 Merging within a wave

Script filename: *Program3-CVACS_Merging_files_into_a_cross-section.do*

This program creates a cross-sectional dataset for a given survey round, by merging together the Unvaccinated, Vaccinated and Derived data files.

9.2.2 Merging the panel together

Script filename: *Program4-CVACS_Merging all the CVACS data into a panel*

This file merges together the Survey 1 and Survey 2 data files.

10 Acknowledgements

We are grateful for the contributions of many individuals and organisations to the design and funding of the study, co-creation of the questionnaire, high-quality data collection, data quality assurance, data analysis, results dissemination and data publication.

The Bill & Melinda Gates Foundation
With special thanks:
John VanderHeide
Tom Black

10.1 Consultative group

Kate Alexander	University of Johannesburg
Zameer Brey	Bill & Melinda Gates Foundation
David Harrison	DG Murray Trust
Saul Johnson	Genesis Analytics
Mosima Mabunda	Discovery Vitality
Ammaarah Martinus	Western Cape Government
Jacqui Miot	HE2RO
Gaurang Tanna	National Department of Health

10.2 Crowdsourcing of survey questions

To the many individuals who provided suggestions for survey questions.

10.3 SALDRU technical team

Cally Ardington
Murray Leibbrandt
Nicola Branson
Reza Daniels
Vimal Ranchhod

10.4 CVACS data team

Victoria Basopo
Morne Hoffman
Julia Tatham
Paidamoyo Bodzo
Nuvika Pillay

10.5 Ask Afrika - Data collection

With particular thanks to:
Sanna Fourie
Jan Lourens
Sarina de Beer
Brownyn Erin Penny

10.6 Survey sampling and weights

Ariane Neethling

10.7 Metadata production, data publication and user support

DataFirst
With particular thanks to:
Lynn Woolfrey
Alison Siljeur

11 Data queries

For data-related queries, please contact:

DataFirst Support
support@data1st.org

12 References

Brophy, T., Branson, N., Daniels, R.C., Leibbrandt, M., Mlatsheni, C., & Woolard, I. (2018). *National Income Dynamics Study panel user manual. Release 2018. Version 1.* Cape Town: Southern Africa Labour and Development Research Unit.

Ingle, K., Brophy, T., Daniels, R.C., 2021. *National Income Dynamics Study – Coronavirus Rapid Mobile Survey (LRES-CRAM) 2020 Panel User Manual. Release February 2021.Version 1*. Cape Town: Southern Africa Labour and Development Research Unit.

Kerr, A., Ardington, C. & Burger, R. (2020). Sample Design and Weighting in the NIDS-CRAM survey. NIDS-CRAM Technical Paper B.

Daniels, R., Ingle, K. and Brophy, T., (2022). Determinants of attrition between Waves 1 and 2 of South Africa's National Income Dynamics Study – Coronavirus Rapid Mobile Survey (NIDS-CRAM). *South African Journal of Economics, [online]* Available at: <https://onlinelibrary.wiley.com/doi/10.1111/saje.12318> [Accessed 28 April 2022].

13 Appendix – CVACS ethics approval

Faculty of Commerce
Private Bag X3, Rondebosch, 7701
2.26 Leslie Commerce Building, Upper Campus
Tel: +27 (0) 21 650 4375/ 5748 Fax: +27 (0) 21 650 4369
E-mail: jacques.rousseau@uct.ac.za
Internet: www.uct.ac.za
@Commerce UCT **UCT Commerce Faculty Office**

09 11 2021

Brendan Maughan-Brown, PhD (UCT); Katherine Eyal, PhD (UCT)

School of Economics

University of Cape Town

REF: REC 2021/11/007

CVACS - COVID-19 Vaccine Survey

We are pleased to inform you that your ethics application has been approved. Unless otherwise specified this ethical clearance is valid until 31-Dec-2022 .

Your clearance may be renewed upon application.

Please be aware that you need to notify the Ethics Committee immediately should any aspect of your study regarding the engagement with participants as approved in this application, change. This may include aspects such as changes to the research design, questionnaires, or choice of participants.

The ongoing ethical conduct throughout the duration of the study remains the responsibility of the principal investigator.

We wish you well for your research.

2021.11.09
09:26:40 +02'00'

Jacques Rousseau
Commerce Research Ethics Chair
University of Cape Town
Commerce Faculty Office
Room 2.26 | Leslie Commerce Building

Office Telephone: +27 (0)21 650 2695 / 4375
Office Fax: +27 (0)21 650 4369
E-mail: jacques.rousseau@uct.ac.za
Website: http://www.commerce.uct.ac.za/com/Ethics-in-Research

"Our Mission is to be an outstanding teaching and research university, educating for life and addressing the challenges facing our society."