



WAVE 3

National Income Dynamics
Study (NIDS) – Coronavirus
Rapid Mobile Survey (CRAM)

Panel User Manual

Kim Ingle - University of Cape Town

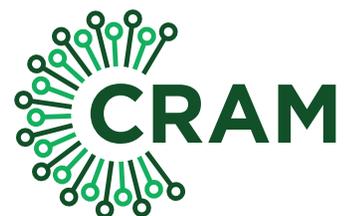
Timothy Brophy - University of Cape Town

Reza Daniels - University of Cape Town

17 February 2021



N.i.D.S.
NATIONAL INCOME DYNAMICS STUDY



CORONAVIRUS RAPID MOBILE SURVEY 2020

Panel User Manual

Kim Ingle, Timothy Brophy and Reza Daniels

Release February 2021

Contents

1. This manual and how to cite it	2
2. Correct citation of the NIDS-CRAM data	2
3. Problematic variables in all waves of NIDS-CRAM	2
3.1. Variables with a high proportion of non-response codes	2
3.1.1. Wave 2	2
3.1.2. Wave 3	3
4. What is NIDS-CRAM?	3
5. Data collection	3
6. Sampling and weights	4
6.1. Sampling	4
6.1.1. Wave 3 top-up sample	4
6.1.2. Attrition between Waves 1, 2 and 3	5
6.2. Weighting	5
6.2.1. Panel weight: Wave 2	5
6.2.2. Wave 3 weights	6
6.2.2.1. Weights for the Original sample only	7
6.2.2.2. Weights for the Original Sample plus Top-up sample	7
6.2.2.3. Weights for the Balanced Waves 1, 2 and 3 panel	8
7. The data	8
7.1. Data files	9
7.1.1. NIDS-CRAM file	9
7.1.2. Derived file	9
7.1.3. Link file	10
7.2. Merging the data within a wave	10
7.3. Non-response and skipped questions in the data	10
7.3.1. Non-response codes	10
7.3.2. Skipped questions	11
7.4. Variable naming convention	11
7.5. Surveyed vs. historical data	11
8. Versioning of the dataset	11
9. Important data limitations	12
9.1. Individual and household level analysis	12
9.2. Regional analysis	12
9.3. Comparing NIDS and NIDS-CRAM	12
9.3.1. Derived and imputed variables	13
9.3.1.1. Geographical variables	13
9.3.1.2. Employment	13
9.3.1.3. Income	13
9.3.2. Occupation	13
9.3.3. Industry	13
9.3.4. Monetary values	13
9.3.5. Inability to match to some NIDS respondents from waves prior to NIDS Wave 5	13
9.3.6. Open-ended questions	14
References	15

1. This manual and how to cite it

This manual is designed to assist users of the NIDS-CRAM data and accompanies the V1.0.0. release of the NIDS-CRAM Wave 3 data. It contains useful information about the survey's design and resulting data for this wave as well as prior waves' data. Users wishing to cite this manual should use the following citation:

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

2. Correct citation of the NIDS-CRAM data

This manual was prepared with the V1.0.0. release of the NIDS-CRAM Wave 3 data. Users wishing to cite this version of the dataset should use the following citation:

National Income Dynamics Study-Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2021, Wave 3 [dataset]. Version 1.0.0. Cape Town: Allan Gray Orbis Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2020. Cape Town: DataFirst [distributor], 2021.

Users wishing to cite the latest version of the Wave 1 or Wave 2 data, should use the following citations respectively:

National Income Dynamics Study-Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2020, Wave 1 [dataset]. Version 1.1.0. Cape Town: Allan Gray Orbis Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2020. Cape Town: DataFirst [distributor], 2020.

National Income Dynamics Study-Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2020, Wave 2 [dataset]. Version 1.1.0. Cape Town: Allan Gray Orbis Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2020. Cape Town: DataFirst [distributor], 2020.

3. Problematic variables in all waves of NIDS-CRAM

3.1. Variables with a high proportion of non-response codes

Certain variables have a high proportion of non-response codes (see section 7.3.1). Researchers are reminded to check the response rate for questions that elicit data they aim to use in their analyses and see whether they think that the response rate is sufficient for their research purposes.

Researchers are thus advised to exercise EXTREME caution if using these variables and to disclose in their research what the response rate for the question was:

3.1.1. Wave 2

Particular attention is drawn to question Bc5 *w2_nc_nocld* in the Wave 2 data where the proportion of -3 "Missing" non-response is notably high due to implementation issues. This question was erroneously skipped over for those where *w2_nc_nou18res* was equal to "0".

3.1.2. Wave 3

Attention is also drawn to question Ba9 *w3_nc_mv2017* in the Wave 3 data, asked of top-up respondents in Wave 3, where the proportion of -3 “Missing” non-response is high for these top-up respondents due to implementation issues.

4. What is NIDS-CRAM?

NIDS-CRAM is the National Income Dynamics Study - Coronavirus Rapid Mobile Survey. The survey investigates the socioeconomic impacts of the national lockdown associated with the State of Disaster declared in South Africa in March 2020, and the social and economic consequences of the global Coronavirus pandemic.

NIDS-CRAM forms part of a broader study called the [Coronavirus Rapid Mobile Survey \(CRAM\)](#) which aims to inform policy using rapid reliable research on income, employment and welfare in South Africa, in the context of the global Coronavirus pandemic. The study is run by researchers from three universities: the [Stellenbosch University, University of Cape Town \(UCT\)](#) and the [University of the Witwatersrand](#). The NIDS-CRAM questionnaire incorporates inputs from members of working and reference groups which comprise researchers from multiple universities and government officials and analysts, among others. The NIDS-CRAM survey data collection and production operations were implemented by the [Southern Africa Labour and Development Research Unit \(SALDRU\)](#) based at UCT’s School of Economics.

NIDS-CRAM is a special follow up with a subsample of adults from households in the [National Income Dynamics Study \(NIDS\)](#) Wave 5 (2017). NIDS is a panel survey which was run by SALDRU. It studies the well-being of South Africans, the households they live in and how these change over time. It started in 2008 when around 28 000 people across South Africa were interviewed. People were then re-interviewed, together with anyone they were living with at the time, every two to three years. Almost 40 000 people were interviewed in 2017. For more information about NIDS, consult the [NIDS website](#) and access the [NIDS data through DataFirst’s Open Data Portal](#).

Though NIDS-CRAM is a follow up with NIDS respondents, in comparison to the core NIDS panel study, NIDS-CRAM uses a much shorter questionnaire, with a focus on the Coronavirus pandemic and the national lockdown. The mode of the survey also changed from face-to-face to Computer Assisted Telephone Interviewing (CATI) surveys which will be repeated over several months.

5. Data collection

CATI interviews were conducted in the preferred official South African language¹ of the respondent.

Table 1 contains the start and end date of the data collection periods for each of the waves.

Table 1: Data collection dates

Wave	Start date	End date
1	7 May 2020	27 June 2020
2	13 July 2020	13 August 2020
3	2 November 2020	13 December 2020

¹ Wave 1 questionnaires were translated into 10 of South Africa’s 11 official languages. The excluded language was Ndebele.

6. Sampling and weights

6.1. Sampling

In NIDS, sample members could either be Continuing Sample Members (CSMs) or Temporary Sample Members (TSMs). CSMs were interviewed in every wave (data collection round) of NIDS, whereas TSMs were interviewed only in the wave(s) when they were co-resident with a CSM. For information regarding the original NIDS Wave 1 sampling, see Leibbrandt et al. (2009). For information on the NIDS Wave 5 top-up in 2017, see Branson (2019).

The sample frame for NIDS-CRAM is the NIDS Wave 5 CSMs and TSMs who were 18 years or older at the time of the NIDS-CRAM Wave 1 fieldwork preparation in April 2020. The sample was drawn using a stratified sampling design. No attempt was made to check whether successfully re-interviewed individuals resided in the same households as they did in Wave 5. Also, individuals from larger households were more likely to be sampled than individuals from smaller households (Kerr et al., 2020). These have important implications for analysis which are discussed in section 9.1.

6.1.1. Wave 3 top-up sample²

Due to attrition between NIDS-CRAM Waves 1 and 2 (roughly 19%), it was decided to replenish the sample by adding randomly selected individuals from the original NIDS Wave 5 sample who were not yet selected for NIDS-CRAM Wave 1. Where strata sizes allowed, the sampling team selected from the different strata inversely proportional to the Wave 1 NIDS-CRAM response rate multiplied by the Wave 2 non-attrition rate. A total of 3078 individuals were selected in this way, and contacted by the fieldworkers for NIDS-CRAM Wave 3. Of these 3078 individuals, 1084 agreed to participate in NIDS-CRAM Wave 3, resulting in the following total sample for Wave 3:

Table 2: NIDS-CRAM sample of the Wave 3 sample members

Sample	Frequency	Percent
Wave 1 NIDS-CRAM Sample	7,073	86.7%
Wave 3 Top-Up Sample	1,084	13.3%
Total	8,157	100%

From Wave 3 onwards, the variable *nids_cram_sample* in the Link file and *w'i'_nc_sample* in the NIDS-CRAM file, shows whether the sample member is from the original NIDS-CRAM Wave 1 sample or the NIDS-CRAM Wave 3 top-up sample.

The final interview outcomes for Wave 3 sample (unweighted) are as follows:

Table 3: Wave 3 interview outcomes

Wave 3 NIDS-CRAM outcome	Frequency	Percent
Successfully Interviewed in Wave 3	6,130	75.2%
Refused/ Not Available	206	2.5%
Not Contacted in Wave 3	1,717	21.1%
Not Tracked in Wave 3	30	0.4%
Moved outside of South Africa (SA) by Wave 3	15	0.2%
Deceased in Wave 3	37	0.5%
Deceased in a prior wave	22	0.3%
Total	8,157	100%

² We are grateful to the sampling team of Cally Ardington, Rulof Burger and Andrew Kerr for assisting us with the writing of this section.

6.1.2. Attrition between Waves 1, 2 and 3

When conducting an analysis of attrition in NIDS-CRAM, it is important to note that it is possible for there to be negative attrition from the third wave onwards. Negative attrition is when a participant participates in Wave 1, drops out (attrits) in a later wave (e.g. Wave 2), and then participates again in a later wave (e.g. Wave 3). A summary of survey participation (including attrition) is provided in the table below.

Table 4: Sample participation and attrition in NIDS-CRAM (Waves 1, 2 and 3)

Sample participation	Frequency	Percent
Participated in all 3 waves (balanced panel)	4,508	55.3%
Participated in Wave 1 and 2, attrite in Wave 3	1,114	13.7%
Participate Wave 1 and 3, attrite in Wave 2 (i.e. negative Wave 2 attrition in Wave 3)	522	6.4%
Participated in Wave 1 only	793	9.7%
Participated in Wave 3 only (top-up sample)	1,084	13.3%
Out of scope (uncontactable, not tracked, deceased, moved outside SA)	136	1.7%
Total	8,157	100%

6.2. Weighting

As Kerr et al. (2020) explain, the weighted NIDS-CRAM survey data reflects the outcomes in 2020 (for Waves 1, 2 or 3) for a broadly representative sample of those 15 years and older from NIDS Wave 5 in 2017 who were followed up 3 years later; implying that the minimum age of the sample is now 18 years old. When interpreting or discussing results from the NIDS-CRAM survey, researchers should note this sample and representivity. For example, when discussing employment losses, a researcher should note: “These results are for a broadly representative sample of South African adults from 2017, who were re-interviewed in 2020 for NIDS-CRAM.”

It is important to remember that NIDS-CRAM is a sample of individuals; therefore, there is statistical uncertainty for any point estimate of variables of interest. Researchers should use the weight, cluster and stratum variables provided with the data in order to correctly estimate parameters of interest. The cluster and stratum variables to be used are the original NIDS cluster and stratum. Variances for some statistics will be higher than for surveys with larger sample sizes like Statistics South Africa’s Quarterly Labour Force Survey or General Household Survey.

In Stata the command to set the complex survey design in NIDS-CRAM Wave 1 would be:

```
svyset cluster [weight = w1_nc_wgt], strata(stratum)
```

This command line should be suitably adjusted for each successive wave of NIDS-CRAM, with the relevant variables. The weights in NIDS-CRAM are the NIDS Wave 5 post-stratified weights, adjusted for the probability of being selected into NIDS-CRAM in 2020, and then adjusted for non-response. The weights are then trimmed to the 1st and 99th percentile. This weighting adjustment process is the same as that followed in NIDS Wave 5. For more detail on the sampling and weights, please refer to Kerr et al. (2020).

6.2.1. Panel weight: Wave 2

Panel surveys invariably suffer from attrition and NIDS-CRAM is no exception. Table 5 shows the Wave 2 outcomes for the 7073 individuals who were successfully interviewed in Wave 1. The attrition rate between the two waves was around 19 percent³.

³ Individuals who have died or emigrated are not considered to have dropped out from the panel for the purposes of estimating non-response.

Table 5: Wave 2 outcomes for Wave 1 sample members

Wave 2 outcome	Frequency	Percent
Successfully completed interview in Wave 2	5676	80.2%
Refused/ Not Available	127	1.8%
Not Contacted in Wave 2	1212	17.1%
Not Tracked in Wave 2	29	0.4%
Moved outside of SA by Wave 2	7	0.1%
Deceased in Wave 2	22	0.3%
Total	7073	100%

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 Wave 2 is to model the probability of being re-interviewed in Wave 2 conditional on a range of Wave 1 characteristics. The following explanatory variables were included in the final probit model – race, sex, age categories (<30, 30-39, 40-50, >50), province⁴, urban location, an indicator that household income was missing, an indicator that the individual was employed at the time of the Wave 1 interview and NIDS Wave 5 household per capita income quartiles. In this multivariate model, attrition was statistically significantly higher among Coloureds and Indians, urban dwellers, the employed, those with missing household income and individuals in the top per capita income quintile in NIDS Wave 5. Re-interview rates were highest for individuals older than 50 years of age.

The second step is to adjust the NIDS-CRAM Wave 1 weights by the inverse of the probability of appearing in Wave 2. Specifically, the panel weights are the product of the NIDS-CRAM Wave 1 weights and the inverse of the conditional probability of responding in Wave 2 estimated from the probit model. The final step is to trim the weights at the 1st and 99th percentiles.

From Wave 2, the name of the wave's weight is w*i*'_nc_pweight, where 'i' refers to the wave number. As a result, in Stata the command to set the complex survey design in NIDS-CRAM Wave 2 would be:

```
svyset cluster [weight = w2_nc_pweight], strata(stratum)
```

6.2.2. Wave 3 weights⁵

Three sets of weights are provided for NIDS-CRAM Wave 3: “Wave 3 weight excluding top-up” if researchers use only the original sample and omit the top-up sample; “Wave 3 weight including top up” if researchers include both the original and the top-up samples; and “Balanced panel weights for waves 1 to 3”. Individuals from the top-up sample only have the second set of weights.

The “Wave 3 weight including top up” should only be used when undertaking cross sectional work using both the top up and original samples together; e.g. when calculating point estimates of states at a point in time, such as when calculating the national employment rate in Wave 3.

The “Wave 3 weight excluding top-up” should be used when analysing the longitudinal sample in NIDS-CRAM, such as when evaluating employment transitions between Wave 1 and Wave 3. To describe individual changes in employment status in Wave 1 requires the group of individuals present in both Waves 1 and 3. Those present in Wave 1 but not Wave 3 have attrited, so the “Wave 3 weight excluding top-up” must be used because it is a weight adjusted for attrition between Waves 1 and 3.

⁴ The NIDS-CRAM Wave 1 province was replaced with the NIDS Wave 5 province for eight individuals who had missing data in NIDS-CRAM Wave 1 on this variable.

⁵ This section is written by Cally Ardington, Rulof Burger and Andrew Kerr. Nicola Branson from SALDRU provided very helpful guidance in thinking through the weights for the top-up sample.

The “Balanced panel weights for Waves 1 to 3” should be used when analysing three period transitions on the balanced panel from all three waves.

There is no weight that only adjusts for attrition between Waves 2 and 3, so if users want to describe changes between Waves 2 and 3 then researchers can either use the “Wave 3 weight excluding top up”, or the balanced panel weight (researchers can also generate their own weight to adjust for attrition between Wave 2 and 3).

6.2.2.1. Weights for the Original sample only

The attrition rate between Wave 1 and Wave 3, for Wave 1 sample members (i.e. excluding the top-up sample), was 28%.

Table 6: Wave 3 outcomes for Wave 1 sample members

Wave 3 outcome	Frequency	Percent
Successfully completed interview in Wave 3	5046	71.3%
Refused/ Not Available	206	2.9%
Not Contacted in Wave 3	1717	24.3%
Not Tracked in Wave 3	30	0.4%
Moved outside of SA by Wave 3	15	0.2%
Deceased in Wave 3	37	0.5%
Deceased in a prior wave	22	0.3%
Total	7073	100%

The weights when excluding the top-up sample are created in a similar manner to the way the NIDS-CRAM Wave 2 weights were created. A probit regression was used to adjust the Wave 1 weights for attrition between Wave 1 and Wave 3. Since there was some negative attrition between Wave 2 and Wave 3 (some individuals who were interviewed in Wave 1, not in Wave 2 but then were interviewed in Wave 3 again; see section 6.1.2), Wave 2 was ignored in calculating the Wave 3 weights. The probit used NIDS Wave 5 characteristics, as was the case for the Wave 1 non-response probit⁶.

The Wave 3 weights without the top-up sample are thus calculated as:

$$\text{Wave 3 weight without top-up} = W_{Nw5\ i} \times \frac{1}{P_{Design_news}} \times \frac{1}{p_{NCw1\ i}} \times \frac{1}{p_{NCw3\ i}}$$

$W_{Nw5\ i}$ is the original NIDS Wave 5 cross section calibrated weight, P_{Design_olds} is the design probability of selection given the sampling method to select the NIDS-CRAM Wave 1 sample (which was the same for all individuals in the stratum), $p_{NCw1\ i}$ is the NIDS-CRAM Wave 1 response probability and $p_{NCw3\ i}$ is the NIDS-CRAM Wave 3 non-attrition probability. i indicates individual and s indicates stratum.

$p_{NCw1\ i}$ was part of the Wave 1 and Wave 2 weights and $p_{NCw3\ i}$ is estimated using a probit on response in Wave 3 conditional on an individual successfully interviewed in Wave 1.

These weights are then trimmed at the 1st and 99th percentile.

6.2.2.2. Weights for the Original Sample plus Top-up sample

The weights when including the top-up sample are a little more complicated. The discussion requires knowledge of the sampling strategy for NIDS-CRAM Wave 1. Please read the sampling document (Kerr et al., 2020) for more detail.

⁶ The Wave 2 weights used some Wave 1 characteristics, but this was not done for Wave 3.

There are two ways of an individual appearing in Wave 3 - either through being selected for Wave 1 (and then being interviewed in Wave 1 and Wave 3), or through being selected in the top-up sample for Wave 3 (see section 6.1.1). But a priori the probability of selection does not depend on whether an individual was selected in Wave 1 or the Wave 3 top-up. Every individual in the same stratum had the same probability of selection - this was the final stratum sampling rate after the Wave 3 top-up sample had been selected.

The weight incorporating the NIDS Wave 5 calibrated weight and the new design weight is then adjusted in two ways, depending on whether the person was selected for the original NIDS-CRAM Wave 1 sample or in the top-up sample in Wave 3. This weight “Wave 3 weight including top-up” is a weight for both the NIDS-CRAM Wave 1 sample members and the NIDS-CRAM Wave 3 top-up sample members. Using these top-up inclusive weights and including both groups together provide a broadly representative sample of the outcomes of the NIDS Wave 5 sample members during Wave 3.

For the original Wave 1 sample, the weight is similar to the weight above, except that the new probability of selection is used:

$$\text{Wave 3 weight including top-up} = W_{Nw5\ i} \times \frac{1}{P_{Design_news}} \times \frac{1}{P_{NCw1\ i}} \times \frac{1}{P_{NCw3\ i}}$$

P_{Design_news} is the design probability of selection after the top-up sampling was performed.

For those selected in the top-up sample the Wave 3 weight that includes this top-up sample must only be adjusted for non-response of this group in Wave 3:

$$\text{Wave 3 weight including top-up} = W_{Nw5\ i} \times \frac{1}{P_{Design_news}} \times \frac{1}{P_{NCw3\ TU}}$$

Where $p_{NCw3\ TU}$ is the Wave 3 response probability for the top-up sample.

The top-up weights are then trimmed for the top-up and the original sample to the 1st and 99th percentile.

6.2.2.3. Weights for the Balanced Waves 1, 2 and 3 panel

The balanced panel weights are used to analyse transitions between Waves 1, 2 and 3 on the sample of individuals who completed the survey in all three waves. The weights are created following a similar approach to that used to create the NIDS-CRAM Wave 2 weights and “Wave 3 weight excluding top-up”. The sampling team now treat anyone who did not complete either the Wave 2 or Wave 3 survey as attriters, and use a probit regression to adjust the Wave 1 weights for this attrition. The probit used NIDS Wave 5 characteristics as predictors of attrition.

$$\text{Balanced panel weight} = W_{Nw5\ i} \times \frac{1}{P_{Design_olds}} \times \frac{1}{P_{NCw1\ i}} \times \frac{1}{P_{NCw2w3\ i}}$$

$W_{Nw5\ i}$, P_{Design_olds} , and $p_{NCw1\ i}$ are defined above, and $p_{NCw2w3\ i}$ is the probability of non-attrition in both Wave 2 and 3.

7. The data

This data release contains three data files. Each has a single record per individual, with the unique identifier *pid*. This identifier for respondents in NIDS-CRAM is the same as their identifier in NIDS. Users wishing to use NIDS-CRAM together with NIDS are referred to section 9.3, which contains some important considerations for this.

7.1. Data files

7.1.1. NIDS-CRAM file

This is the main data file containing data from responses to the questions in the questionnaire.

From Wave 2 onwards, it only includes those that are believed to be alive at the time of that wave. Apart from those successfully interviewed, this data file includes those with all reasons for non-response, other than death. Researchers are thus reminded to control for interview outcome by making use of the *w`i`_nc_outcome* variable.

Since Stata limits variable labels to a length of 80 characters, a shortened version of the question is used for the variable labels in this data file. This data file does not contain any variables which are potentially disclosive. As a result, there are a few questions which appear in the questionnaire for which there is no corresponding data in the public release version.

For some questions, the pre-defined answer categories were read out to the respondent, while for others, they were not and the enumerators were instructed to select the appropriate answer option(s). Please consult the questionnaire to view the questions as they were asked.

Where a question with categorical responses included an “Other (specify)” option, the answers to this option were reviewed and where possible coded into an existing category. Where there were enough similar responses which did not fall into an existing category, new categories were added to group these responses for ease of analysis. These have also been added to the questionnaire for the user’s reference.

Occupation data coded according to the International Standard Classification of Occupations (ISCO) 1-digit level are included in this data file. ISCO-08 codes are used. These are coded from the respondent’s answers to the occupation questions. Duration of interview and date of interview are also included.

In Wave 2 industry codes for employment are also included in this data file. These codes link the main goods or services provided by the employer to the industry description. The industry codes used are those found in Statistics South Africa’s General Household Survey (2005) industry code list. These codes were truncated to the one-digit level and included in the public release data.

From Wave 3 onwards, this file contains a variable *w`i`_nc_sample* which shows whether the sample member is from the original NIDS-CRAM Wave 1 sample or the NIDS-CRAM Wave 3 top-up sample.

7.1.2. Derived file

This data file contains derived variables which are not collected with questions asked directly of the respondent but are calculated or imputed from their responses to questions and other information. The derived variables include:

- Weights
 - The weight discussed in section 6.2.
- Best estimates of some demographic characteristics of the respondent
 - These are best estimates of the characteristics considering both NIDS panel and NIDS-CRAM responses.
- Age intervals
 - Age intervals are calculated from the respondent’s best age.
- National lockdown level at the time of interview

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

7.1.3. Link file

This data file contains interview outcomes for each respondent in the different NIDS-CRAM waves, together with their NIDS Wave 5 individual interview outcome. 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 NIDS Wave 5 2017 top-up, the variable *sample* is included in the Link file. This variable identifies which sample individual respondents originated from. It takes on the value 1 for “2008 sample” and 2 for “2017 sample”. Similarly, from Wave 3 onwards, this file contains a variable *w`i`_nc_sample* which shows whether the sample member is from the original NIDS-CRAM Wave 1 sample or the NIDS-CRAM Wave 3 top-up sample.

The cluster and strata variables which are discussed in section 6.2 are included in the Link file.

From Wave 3 onwards, the variable *nc_wave_died* is included in the link file. For respondents who died during NIDS-CRAM, this variable contains which wave their death was first recorded.

7.2. Merging the data within a wave

In order to merge observations for individuals from the three data files, users should merge on the person level identifier *pid*. The relevant Stata code for merging V1.0.0. of the Wave 3 data is as follows:

```
use "Link_File_NIDS-CRAM_Wave3_Annon_V1.0.0.dta", clear
merge 1:1 pid using "NIDS-CRAM_Wave3_Annon_V1.0.0.dta"
drop _merge
merge 1:1 pid using "derived_NIDS-CRAM_Wave3_Annon_V1.0.0.dta"
drop _merge
```

7.3. Non-response and skipped questions in the data

7.3.1. Non-response codes

NIDS-CRAM uses the same non-response codes as were used in NIDS. In general, these are as follows:

Table 7: Typical non-response codes in NIDS-CRAM

	General	Month	Year
Missing	-3	33	3333
Not applicable	-5	55	5555
Refused	-8	88	8888
Don't know	-9	99	9999

-3 “Missing” is a data quality code which is only used post-fieldwork. It is applied to show the 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.

7.3.2. Skipped questions

As a result of the questionnaire design, some questions are not asked of all respondents, as they are skipped for some. The “Skips” column of the questionnaire shows where skips are implied by answers to certain questions. When a question has been skipped due to questionnaire design, the following will apply:

- For a non-string variable, this will be denoted by a “.”, or system missing.
- For string variables, the string will be empty.

7.4. Variable naming convention

Variables that are wave specific have the prefix “*w*i*_nc*”, where *i* is the wave number.

7.5. Surveyed vs. historical data

From Wave 2 onwards, certain questions were not re-asked of some respondents. This was done to avoid re-asking respondents time-invariant data that we have collected previously. Where this was the case, the historic data was used in the NIDS-CRAM data file. In order for users to differentiate between this historical data and the data which was collected in that wave, flag variables have been created. An example of this is *w2_nc_edschgrd_flg*.

8. Versioning of the dataset

Much like NIDS Wave 5, the versioning of NIDS-CRAM aims to inform users of major, minor and patch fix releases of the data to ensure they use the most up-to-date data for their research. As explained by Brophy et al. (2018), the versioning format can be broken down as follows:

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

Major releases

- Refers to the number of a major release. A major release only occurs when a new wave is added to the panel.
- e.g. V1.0.0, here 1 indicates the first major release of the dataset.

Minor releases

- Refers to the number of a minor release. A minor release is the release of a single wave/s that occurs between major releases, i.e. a release of corrected data where no new wave is added to the panel.

- 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) which are created to correct specific errors in the data. The scripts are given to users to run on the existing data 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.

9. Important data limitations

9.1. Individual and household level analysis

NIDS-CRAM sampled individuals from NIDS Wave 5. Unlike previous waves of NIDS, NIDS-CRAM did not attempt to interview or collect information on everyone currently living with the sampled individual. This change in sampling protocol was carefully considered, taking into account the main goals and constraints of NIDS-CRAM. The limits of telephonic surveys with respect to questionnaire length and complexity was a key factor. No attempt was made to check whether successfully re-interviewed individuals resided in the same households as they did in NIDS Wave 5. Also, individuals from larger households are more likely to be sampled than individuals from smaller households. Researchers can therefore not use NIDS-CRAM to conduct household-level analysis. However, it is possible to estimate statistics at an individual level about household living conditions. It would be legitimate to state “For a broadly representative sample of adults from 2017, who were re-interviewed in 2020, we estimate that X% of adults live in households receiving a government grant”. However, it is not legitimate to estimate that Y% of households received a government grant. In the same way one could say that Z% of adults live in households where children went hungry, but not that A% of households had children going hungry or that B% of children went hungry (Kerr et al., 2020).

9.2. Regional analysis

As NIDS was only designed to be nationally, not provincially representative, we do not recommend using NIDS-CRAM or NIDS to calculate provincial totals.

9.3. Comparing NIDS and NIDS-CRAM

Though the NIDS-CRAM respondents can be matched to their observations in the core NIDS panel, using their unique identifier *pid*, researchers should exercise caution when making comparisons between these two datasets. Though the NIDS-CRAM sample members were drawn from the NIDS Wave 5 sample, NIDS-CRAM and NIDS differ in important methodological ways:

- Mode differences:
 - In-person interviews in NIDS and CATI in NIDS-CRAM.
- Questionnaire type differences:
 - Household level questions: NIDS asked the oldest women or most knowledgeable person in the household (where possible) about the household and its members through a separate household questionnaire. NIDS-CRAM asks the selected respondent questions about their households.
 - Since NIDS interviewed all household residents and had a household level questionnaire, it allowed for cross-individual and household-to-individual questionnaire validation checks, whereas NIDS-CRAM is based on one questionnaire.
- Wording differences that imply definition differences.
- Fewer and less detailed questions in NIDS-CRAM than in NIDS (due to the mode change requiring a shorter questionnaire):

- o Some variables were calculable from answers in NIDS, whereas NIDS-CRAM needed to ask these directly, given the limits in questionnaire length. An example would be the NIDS collection of a full household roster, which enables researchers to calculate the number of household members (and residents) of various ages, whereas NIDS-CRAM asks the selected respondent directly for the number of household members of specific ages.

A few examples of specific considerations to bear in mind when comparing NIDS and NIDS-CRAM follow. Please note that this list is not exhaustive, and researchers are encouraged to carefully consider whether it is appropriate to compare specific data collected in NIDS and NIDS-CRAM.

9.3.1. Derived and imputed variables

9.3.1.1. Geographical variables

NIDS respondents' geographical information was derived from the GPS coordinates which were collected at their dwelling when they were interviewed. Conversely, all NIDS-CRAM derived geographic information is based on the respondent's self-reported area that they live in.

Following Wave 1, attempts were made to collect more detailed geographic data from Wave 2 onwards, by improving the drop-down lists that the interviewers selected from when capturing respondents' reported location.

9.3.1.2. Employment

The employment status variable is derived in NIDS-CRAM using a similar method to how it was derived in NIDS. However, the questions which are used are different and less detailed in NIDS-CRAM. Further, some of the questions on employment differ between NIDS and NIDS-CRAM; for example, the time period referenced in the question.

9.3.1.3. Income

The level of detail of the questions in NIDS-CRAM does not allow for full-imputations of income as in NIDS.

9.3.2. Occupation

In NIDS respondents were asked about the type of job that they do as well as the tasks that they perform. In NIDS-CRAM no question was asked on the tasks people perform. In NIDS-CRAM the occupation code is coded based solely on the job type, rather than job type and tasks, as in NIDS.

9.3.3. Industry

Questions used to derive industry of employment were only asked from Wave 2. Industry is thus not available for Wave 1 analysis.

9.3.4. Monetary values

When comparing monetary values over time, such as in earlier NIDS waves, 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.

9.3.5. Inability to match to some NIDS respondents from waves prior to NIDS Wave 5

The sample frame for NIDS-CRAM is the NIDS Wave 5 CSMs and TSMs who were 18 years or older at

the time of the NIDS-CRAM Wave 1 fieldwork preparation in April 2020. Since NIDS Wave 5 included a sample top-up and Wave 5 TSMs were not interviewed throughout the NIDS panel, merging NIDS-CRAM with NIDS waves prior to Wave 5, will result in inability to match some respondents from NIDS-CRAM. Using a panel balanced with waves prior to Wave 5 of NIDS would thus present representivity issues for the researcher to consider. Further, the researcher would need to calculate the relevant panel weights for this analysis themselves.

9.3.6. Open-ended questions

The NIDS-CRAM questionnaire includes several open-ended questions where the enumerator either captures the response as falling within a pre-existing category, or records their translation of the response to be coded into a category later. 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.

References

Branson, N. (2019). Adding a Top-Up Sample to the National Income Dynamics Study in South Africa. NIDS Technical Paper 8.

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.

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

Leibbrandt, M., Woolard, I. & de Villiers, L. (2009). Methodology: Report on NIDS Wave 1. NIDS Technical Paper 1.

For further information please see cramsurvey.org and nids.uct.ac.za