Panel User Manual

WAVE 5

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

Kim Ingle - University of Cape Town

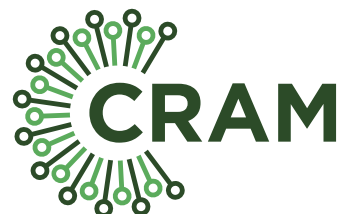**Timothy Brophy** - University of Cape Town

Reza Daniels - University of Cape Town

8 July 2021

Panel User Manual

Kim Ingle, Timothy Brophy and Reza Daniels

Release July 2021

Contents

1. This manual and how to cite it

This manual is designed to assist users of the NIDS-CRAM data and accompanies the version 1.0.0. release of the NIDS-CRAM Wave 5 data. It contains useful information about the survey's design and resulting data for this wave as well as prior waves' (data collection round's) 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 - 2021 Panel User Manual. Release July 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 version 1.0.0. release of the NIDS-CRAM Wave 5 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 5 [dataset]. Version 1.0.0. Cape Town: FEM Education Foundation and Michael and Susan Dell Foundation [funding agencies]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2021. Cape Town: DataFirst [distributor], 2021. DOI: https://doi.org/10.25828/awhe-t852

Users wishing to cite the latest version of the Wave 1, Wave 2, Wave 3 or Wave 4 data, should use the following citations respectively:

National Income Dynamics Study-Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2020, Wave 1 [dataset]. Version 3.0.0. Cape Town: Allan Gray Orbis Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2020. Cape Town: DataFirst [distributor], 2020. DOI: https://doi.org/10.25828/7tn9-1998

National Income Dynamics Study-Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2020, Wave 2 [dataset]. Version 3.0.0. Cape Town: Allan Gray Orbis Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2020. Cape Town: DataFirst [distributor], 2020. DOI: https://doi.org/10.25828/5z2w-7678

National Income Dynamics Study-Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2020, Wave 3 [dataset]. Version 3.0.0. Cape Town: Allan Gray Orbis Foundation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2020. Cape Town: DataFirst [distributor], 2021. DOI: https://doi.org/10.25828/s82x-nx07

National Income Dynamics Study-Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2021, Wave 4 [dataset]. Version 2.0.0. Cape Town: FEM Education Foundation and Michael and Susan Dell Foundation [funding agencies]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2021. Cape Town: DataFirst [distributor], 2021. DOI: https://doi.org/10.25828/y5qj-x095

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

5.1. Method of data collection

CATI interviews were conducted in the preferred official South African language[1] of the respondent.

5.2. Data collection dates

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
4	2 February 2021	10 March 2021
5	6 April 2021	11 May 2021

5.3. Contact attempts

5.3.1. Call attempts

Phone numbers for each respondent were provided to the fieldwork company at the start of each wave. These phone numbers could belong to either the respondent or someone who might know how to contact the respondent. The number of phone numbers provided per respondent varied over the waves. This was due to feedback on disconnected phone numbers, newly collected phone numbers in prior waves, and in some cases limits as a result of operational and time constraints.

The service provider needed to attempt all phone numbers they were provided with at the beginning of the wave, as well as any new numbers that they were provided by respondent contacts during the wave. Each phone number needed to be attempted a number of times before the service provider could stop attempting to make contact with the respondent and mark them as "Not Contacted" in that wave. The derivation of call attempt rules was an iterative process over the panel, informed by the success rates of increased attempts of different contact types, as well as operational feasibility.

5.3.2. SMSs

Each wave before respondents were called, they were sent an SMS. This SMS varied over the waves, but different iterations included that the respondent will be called in the coming days/weeks; the airtime incentive; the lucky draw and thanks for participating. These SMSs were either sent in batches or all in one go, depending on the speed at which portions of the sample were first going to be contacted in a wave.

When operational feasibility allowed for additional call attempts, sample members who were proving difficult to get a hold of (and could qualified or were about to qualify as "Not Contacted"), were sent an additional reminder SMS by the service provider, 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 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 that wave.

---

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

5.4. Non-fielded sample members

From Wave 2 onwards sample members who had permanently refused to be a part of the survey in a previous wave, as well as those who had died, were not reattempted.

By Wave 4, a portion of the original Wave 1 sample had refused in both Waves 2 and 3 or had been "Not Contacted" in both Waves 2 and 3. Since these respondents were highly unlikely to be successfully interviewed in Wave 4 or Wave 5, they were not reattempted from Wave 4. These sample members make up the majority of those who have the outcome "Not Tracked" in Wave 4. The remainder of those with the outcome "Not Tracked" in a wave, were those who were not successfully tracked in that wave, due to implementation issues.

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 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 8.1.

In Wave 1 of NIDS-CRAM, a total of 7,073 respondents were successfully interviewed.

6.1.1. Wave 3 top-up sample[2]

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: Composition of NIDS-CRAM Wave 3 sample who participated in at least one wave

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* (where {i} denotes the wave number) 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.

---

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.

The final interview outcomes for the total Wave 3, 4 and 5 samples (unweighted) are as follows:

Table 3: Wave 3, Wave 4 and Wave 5 interview outcomes

Outcome	Wave 3		Wave 4		Wave 5	
	Freq.	Percent	Freq.	Percent	Freq.	Percent
Successfully interviewed in this wave	6,130	75.2%	5,629	69.0%	5,862	71.9%
Refused/ not available in this wave	206	2.5%	176	2.2%	210	2.6%
Not contacted in this wave	1,717	21.1%	1,561	19.1%	1,256	15.4%
Not tracked in this wave	30	0.4%	694	8.5%	708	8.7%
Moved outside of South Africa (SA) in this wave	15	0.2%	5	0.1%	15	0.2%
Deceased in this wave	37	0.5%	33	0.4%	14	0.2%
Deceased in a prior wave	22	0.3%	59	0.7%	92	1.1%
Total	8,157	100%	8,157	100%	8.157	100%

6.1.2. Sample participation and non-participation over the waves

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). It is also possible for this to happen multiple times over 5 waves. In the table below, we show all possible permutations of survey participation and attrition in NIDS-CRAM, for both the original and top-up samples. In column 1, row 2, we see the balanced panel across all five waves, excluding the top-up sample. This is denoted "w12345". In row three, we see respondents who participated in Waves 1 to 4, then attrited in Wave 5, denoted "w1234,a5".

Table 4: Sample participation and non-participation in NIDS-CRAM Waves 1, 2, 3, 4 & 5

Sample participation in NIDS-CRAM	Frequency	Percent
1. w12345: balanced panel (excluding top-up)	3,456	42.4%
2. w1234,a5: participated in 4 waves	364	4.5%
3. w1235,a4: participated in 4 waves	431	5.3%
4. w1245,a3: participated in 4 waves	450	5.5%
5. w1345,a2: participated in 4 waves	275	3.4%
6. w123,a45: participated in 3 waves	215	2.6%
7. w124,a35: participated in 3 waves	123	1.5%
8. w134,a25: participated in 3 waves	70	0.9%
9. w125,a34: participated in 3 waves	236	2.9%
10. w135,a24: participated in 3 waves	106	1.3%
11. w145,a23: participated in 3 waves	22	0.3%
12. w12,a345: participated in 2 waves	315	3.9%
13. w13,a245: participated in 2 waves	79	1.0%
14. w14,a235: participated in 2 waves	11	0.1%
15. w15,a234: participated in 2 waves	20	0.3%
16. w1,a2345: participated in 1 wave	85	1.0%
17. t345: top-up balanced w345	746	9.2%
18. t34,5: top-up w3&4, a5	88	1.1%
19. t35,4: top-up w3&5, a4	120	1.5%
20. t3,a45: top-up w3; a45	116	1.4%
21. Out of scope in w5	829	10.2%
Total	8,157	100%

6.2. Weighting[3]

As Kerr et al. (2020) explain, the weighted NIDS-CRAM survey data reflects the outcomes in 2020 (for Waves 1, 2 and 3) or 2021 (for Waves 4 and 5) 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 in 2020 was 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."

6.2.1. Wave 1, Wave 2 and Wave 3 weights with and without the suffix "_s"

The weights for Waves 1, 2 and 3 are provided in two forms: with a suffix "_s" and without this suffix. Apart from the balanced panel weight, those with the suffix differ from those without the suffix in that they have been scaled to the NIDS Wave 5 population total and the probit regressions used to make non-response and attrition adjustments for these weights were augmented to be more consistent across different waves. In the case of the balanced panel weights with the suffix "_s", while the final balanced panel weight itself is not scaled, one of the components of the balanced panel weights is the Wave 1 weight, which is scaled in this version of the balanced panel weight. For information on the scaling, see section 6.2.5.

It is recommended that those with the suffix "_s" are used by researchers. However, those without the suffix "_s" have also remained in the data for completeness, as they were the weights available in prior releases of the Waves 1, Wave 2 and Wave 3 data, and could thus be used by those wishing to replicate research results which used these original weights.

6.2.2. Wave 1 weights and the cluster and stratum variables

The weights included in the Wave 1 dataset are contained in the below table. Please refer to section 6.2.1 for why both these weights are included and the differences between them:

Table 5: Weights in Wave 1

Variable name	Full description
w1_nc_wgt	Design weights adjusted for non-response and trimmed.
w1_nc_pweight_s	Weights for Wave 1 adjusted for non-response, trimmed and scaled to NIDS Wave 5 population total.

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_pweight_s], 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 Wave 1 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

---

3   This section is largely written by the NIDS-CRAM sampling team comprising of 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.

weights are then trimmed to the 1st and 99th percentile. This weighting adjustment process is the same as that followed in NIDS Wave 5. The Wave 1 scaled weight *w1_nc_pweight_s* is then also scaled to the NIDS Wave 5 population. For more detail on the sampling and original weights, please refer to Kerr et al. (2020).

The Wave 1 weights are calculated as:

Wave 1 weight = $w_{Nw5\,i} \times \dfrac{1}{P_{Design\_old\,s}} \times \dfrac{1}{p_{NCw1\,i}}$

$W_{Nw5\,i}$ is the original NIDS Wave 5 cross section calibrated weight, $P_{Design\_old\,s}$ is the design probability[4] 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. i indicates individual and s indicates stratum.

The probit regression used to estimate the response probability includes as regressors the NIDS-CRAM sample strata, as well as the NIDS Wave 5 variables for race, gender, language, log household income per capita, whether an urban dweller, province, employment status, household size and the individual outcome.

Both the original Wave 1 weight, *w1_nc_wgt*, and the scaled Wave 1 weight, *w1_nc_pweight_s* were then trimmed to the 1st and 99th percentile. The scaled weights were also scaled up to the NIDS Wave 5 population totals minus deaths and emigration between NIDS Wave 5 and NIDS-CRAM Wave 1 (see section 6.2.5 for more information on this scaling).

6.2.3. Wave 2 weights

Panel surveys invariably suffer from attrition and NIDS-CRAM is no exception. Table 6 shows the Wave 2 outcomes for the 7,073 individuals who were successfully interviewed in Wave 1. The attrition rate between the two waves was around 19 percent[5].

Table 6: Wave 2 outcomes for Wave 1 sample members

Wave 2 outcome	Frequency	Percent
Successfully completed interview in Wave 2	5,676	80.2%
Refused/ not available	127	1.8%
Not contacted in Wave 2	1,212	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	7,073	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 weights included in the Wave 2 dataset are contained in the below table. Please refer to section 6.2.1 for why both these weights are included and the differences between them:

---

4 $P_{Design\_old\,s}$ refers to the design probability of selection for the original sample. This is different from the design probability of selection when including the Wave 3 top-up sample, which is referred to as $P_{Design\_new\,s}$.

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

Variable name	Full description
w2_nc_pweight	Panel weights NIDS-CRAM Wave 1 to Wave 2: Adjusted for attrition and trimmed.
w2_nc_pweight_s	Panel weights NIDS-CRAM Wave 1 to Wave 2: Adjusted for attrition, trimmed and scaled to the NIDS Wave 5 population total.

The weight for Wave 2 is calculated as:

$$\text{Wave 2 weight} = w_{Nw5\,i} \times \frac{1}{P_{Design\_old\,s}} \times \frac{1}{p_{NCw1\,i}} \times \frac{1}{p_{NCw1w2\,i}}$$

$w_{Nw5\,i}$, $P_{Design\_old\,s}$, $p_{NCw1\,i}$ were all part of the Wave 1 weight calculation; where *i* also indicates individual and *s* indicates stratum. $p_{NCw1w2\,i}$ is the NIDS-CRAM Wave 2 response (non-attrition) probability.

The first step in creating the panel weights for Wave 2 is to model the probability of being re-interviewed in Wave 2. For the original weights, the NIDS-CRAM Wave 1 characteristics and some NIDS Wave 5 characteristics were used to do so. The following explanatory variables were included in the final probit model – race, sex, age categories (<30, 30-39, 40-50, >50), province[6], 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.

For the scaled weights only NIDS Wave 5 characteristics were used to estimate the probability of being re-interviewed (i.e. no information from NIDS-CRAM Wave 1 was used). These included race, sex, home language, household income in NIDS Wave 5, urban location, province, employment status, household size and whether or not an individual was successfully interviewed in Wave 5 (only individuals whose households were successfully interviewed in NIDS Wave 5 were in the NIDS-CRAM sample frame but some individuals in those households did not successfully complete an individual questionnaire), and NIDS-CRAM stratum dummies.

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 described above.

The weights were then trimmed to the 1st and 99th percentile. The scaled weights were then rescaled so that the Wave 2 population total matched the NIDS Wave 5 population total minus deaths and emigration (see section 6.2.5 below for more details on rescaling).

6.2.4. Wave 3, 4 and 5 weights

6.2.4.1. Wave 3 weights overview

Three broad 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

---

6   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.

panel weights for Waves 1 to 3". Individuals from the top-up sample only have the second set of weights. These weights are then also provided in two forms, with a suffix "_s" and without this suffix. Refer to section 6.2.1 for why both these types are included and the differences between them.

These weights' variable names and descriptions are shown in the table below:

Table 8: Weights in Wave 3

Variable name	Full description
w3_nc_pweight_extu	Panel weights NIDS-CRAM Wave 1 to Wave 3 for the sample that excludes top-up members (i.e. original Wave 1 sample only): Adjusted for attrition and trimmed.
w3_nc_pweight	Panel weights NIDS-CRAM Wave 1 to Wave 3 for the sample that includes top-up members: Adjusted for attrition and trimmed.
w3_nc_bp_pweight	Balanced panel weights NIDS-CRAM Wave 1, Wave 2 and Wave 3: Adjusted for attrition and trimmed.
w3_nc_pweight_extu_s	Panel weights NIDS-CRAM Wave 1 to Wave 3 for the sample that excludes top-up members (i.e. original Wave 1 sample only): Adjusted for attrition, trimmed and scaled to the NIDS Wave 5 population total.
w3_nc_pweight_s	Panel weights NIDS-CRAM Wave 1 to Wave 3 for the sample that includes top-up members: Adjusted for attrition, trimmed and scaled to the NIDS Wave 5 population total.
w3_nc_bp_pweight_s	Balanced panel weights NIDS-CRAM Wave 1, Wave 2 and Wave 3: Adjusted for attrition, trimmed (Wave 1 input weight scaled to the NIDS Wave 5 population total).

The "Wave 3 weights 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 weights 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.

- The "Balanced panel weights for Waves 1 to 3" should be used when analysing three period transitions on the balanced panel from all three of the Waves 1, 2 and 3.

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 weights for Waves 1 to 3". Researchers can also generate their own weight to adjust for attrition between Wave 2 and 3.

6.2.4.2. Wave 4 and Wave 5 weights overview

The weights in Waves 4 and 5 are as follows:

Table 9: Weights in Wave 4

Variable name	Full description
w4_nc_pweight_extu_s	Panel weights NIDS-CRAM Wave 1 to Wave 4 for the sample that excludes top-up members (i.e. original Wave 1 sample only): Adjusted for attrition, trimmed and scaled to NIDS Wave 5 population total.
w4_nc _pweight_s	Panel weights NIDS-CRAM Wave 1 to Wave 4 for the sample that includes top-up members: Adjusted for attrition, trimmed and scaled to NIDS Wave 5 population total.
w4_nc _bp_pweight_s	Balanced panel weights NIDS-CRAM Wave 1, Wave 2, Wave 3 and Wave 4: Adjusted for attrition, trimmed (Wave 1 input weight scaled to NIDS Wave 5 population total).

Table 10: Weights in Wave 5

Variable name	Full description
w5_nc_pweight_extu_s	Panel weights NIDS-CRAM Wave 1 to Wave 5 for the sample that excludes top-up members (i.e. original Wave 1 sample only): Adjusted for attrition, trimmed and scaled to NIDS Wave 5 population total.
w5_nc _pweight_s	Panel weights NIDS-CRAM Wave 1 to Wave 5 for the sample that includes top-up members: Adjusted for attrition, trimmed and scaled to NIDS Wave 5 population total.
w5_nc _bp_pweight_s	Balanced panel weights NIDS-CRAM Wave 1, Wave 2, Wave 3, Wave 4 and Wave 5: Adjusted for attrition, trimmed (Wave 1 input weight scaled to NIDS Wave 5 population total).

These weights should be used in the following situations:

The "Wave k weight top-up (scaled to Wave 5 of NIDS)" (where k refers to the wave number 4 or 5) should be used when undertaking cross sectional work using both the top-up and original samples together, and can also be used when using the panel element of the original plus top-up sample for Waves 3 and k only.

The "Wave k weight excluding top-up (scaled to Wave 5 of NIDS)" should be used when analysing the longitudinal sample in NIDS-CRAM, such as when evaluating employment transitions between Wave 1 and Wave k.

The "Balanced panel weights for Waves 1 to k (Wave 1 input weight scaled to Wave 5 of NIDS)" should be used when analysing four period transitions on the balanced panel from all waves up to, and including, Wave k.

6.2.4.3. 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%. The attrition rate between Wave 1 and Wave 4 was 31%, and between Wave 1 and Wave 5, it was 28%.

Table 11: Wave 3, 4 and 5 outcomes for Wave 1 sample members only

Outcome	Wave 3		Wave 4		Wave 5	
	Freq.	Percent	Freq.	Percent	Freq.	Percent
Successfully completed interview in this wave	5,046	71.3%	4,792	67.8%	4,996	70,6%
Refused/ not available in this wave	206	2.9%	159	2.3%	184	2,6%
Not contacted in this wave	1,717	24.3%	1,337	18.9%	1,078	15,2%
Not tracked in this wave[7]	30	0.4%	692	9.8%	702	9,9%
Moved outside of SA in this wave	15	0.2%	5	0.1%	15	0,2%
Deceased in this wave	37	0.5%	29	0.4%	10	0,14%
Deceased in a prior wave	22	0.3%	59	0.8%	88	1,2%
Total	7,073	100%	7,073	100%	7,073	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 j (where *j* refers to the wave number 3, 4 or 5). Since there was some negative attrition between Waves 2, 3, 4 and 5 (some individuals who were interviewed in Wave 1, not in a later wave, but then were interviewed in Wave j again; see section 6.1.2), the wave(s) between Wave 1 and Wave j was (were) ignored in calculating the Wave j weights. The probit used NIDS Wave 5 characteristics, as was the case for the Wave 1 non-response probit[8].

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

Wave j weight without top-up $= w_{Nw5\,i} \times \dfrac{1}{P_{Design\_old\,s}} \times \dfrac{1}{p_{NCw1\,i}} \times \dfrac{1}{p_{NCw1wji}}$

$W_{Nw5\,i}$, $P_{Design\_old\,s}$, $p_{NCw1\,i}$ were all part of the Wave 1 and 2 weight calculations; where *i* also indicates individual and *s* indicates stratum. $p_{NCw1wji}$ is the NIDS-CRAM Wave j response (non-attrition) probability between Wave 1 and Wave j.

The probit regressions used to estimate the non-attrition probabilities between Wave 1 and 3, between Waves 1 and 4, and between Waves 1 and 5 include all the same regressors as the non-attrition probit between Waves 1 and 2.

The weights are then trimmed at the 1ˢᵗ and 99ᵗʰ percentile. The Wave 3 scaled weight *w3_nc_pweight_extu_s,* Wave 4 weight *w4_nc_pweight_extu_s* and the Wave 5 weight *w5_nc_pweight_extu_s* are then also rescaled to the NIDS Wave 5 total population minus deaths and emigration.

6.2.4.4. 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.

There are two ways of an individual appearing in a wave from Wave 3 onwards – either through being selected for Wave 1 (and then being interviewed in Wave 1 and Wave j), or through being selected

---

7   Refer to section 5.4 for an explanation of this category of sample members and why this is higher from Wave 4 onwards.

8   The original Wave 2 weights used some Wave 1 characteristics, but to be consistent, the scaled weights all use NIDS Wave 5 characteristics only.

in the top-up sample for Wave 3 (and being interviewed in Wave j, 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.

In Wave 3, the weight incorporating the NIDS Wave 5 calibrated weight and the new design weight was 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.

For the original Wave 1 sample, the weight is similar to the Wave 1 and 2 weights, except that the final stratum sampling rate after the Wave 3 top-up sample had been selected is used:

Wave 3 weight including top-up = $w_{Nw5\,i} \times \dfrac{1}{P_{Design\_new\,s}} \times \dfrac{1}{p_{NCw1\,i}} \times \dfrac{'1}{p_{NCw1w3\,i}}$

$P_{Design\_new\,s}$ 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:

Wave 3 weight including top-up = $w_{Nw5\,i} \times \dfrac{1}{P_{Design\_new\,s}} \times \dfrac{1}{p_{NCw3i\_TU}}$

$p_{NCw3_{TU}}$ is the Wave 3 response probability for the top-up sample.

The Wave k weight (where k refers to the wave number 4 or 5) for the top-up plus original sample incorporates the final stratum sampling rate after the Wave 3 top-up sample had been selected. The design weight is then adjusted differently for the top-up and original sample members.

For the individuals selected in the original sample a probit predicting non-attrition between Wave 1 and Wave k is run, with the same variables as the non-attrition probits for Waves 2 and 3. For individuals selected in the top-up sample a probit predicting non-attrition between Wave 3 and Wave k is run. This has very similar variables included, although some uncommon values of the language and employment status variables are grouped together to avoid the probit predicting attrition perfectly.

For the original Wave 1 sample, the weight is then:

Wave k weight including top-up = $w_{Nw5\,i} \times \dfrac{1}{P_{Design\_new\,s}} \times \dfrac{1}{p_{NCw1\,i}} \times \dfrac{1}{p_{NCw1wk\,i}}$

$P_{Design\_new\,s}$ is again the design probability of selection after the top-up sampling was performed. $p_{NCw1wk\,i}$ is the probability that an individual selected for NIDS-CRAM in the original Wave 1 sample is successfully interviewed in Wave k.

For the sample selected in the Wave 3 top-up, the Wave k weight is:

Wave k weight including top-up = $w_{Nw5\,i} \times \dfrac{1}{P_{Design\_new\,s}} \times \dfrac{1}{p_{NCw3\,iTU}} \times \dfrac{1}{p_{NCw3wk\,TU\,i}}$

$p_{NCw3\,iTU}$ is the probability that an individual selected for the top-up sample responded in Wave 3 and $p_{NCw3wk\,TU\,i}$ is the probability that a top-up sample member that responded in Wave 3 also responded in Wave k.

In Wave 3, 4 and 5 the top-up weights are then trimmed to the 1st and 99th percentile of the weights. The scaled Wave 3 weight *w3_nc_pweight_s*, the Wave 4 weight *w4_nc_pweight_s* and the Wave 5 weight *w5_nc_pweight_s* are then also rescaled to the NIDS Wave 5 population minus deaths and emigration.

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 the respective wave. The "Wave k weight including the top-up sample" can also be used when using the panel element of the original plus top-up sample for Waves 3 and k only, e.g. examining changes in employment status between Wave 3 and Wave k for the combined original plus top-up sample. This should give similar results as when using the Wave 3 and Wave k original sample members only.

6.2.4.5. Weights for the balanced panel

The balanced panel weights released with each Wave j of data from Wave 3 onwards, are used to analyse transitions between all waves up to Wave j on the sample of individuals who completed the survey in all waves up to Wave j.

The weights are created following a similar approach to that used to create the NIDS-CRAM Wave 2 weights and later waves' weights which exclude the top-up. The sampling team now treat anyone who did not complete any of the waves between Wave 2 until Wave j 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.

The balanced panel weight for Wave 5 for example, is thus calculated as follows:

Wave 5 balanced panel weight $= w_{Nw5\,i} \times \dfrac{1}{P_{Design\_old\,s}} \times \dfrac{1}{p_{NCw1\,i}} \times \dfrac{1}{p_{NCw2w3w4w5\,i}}$

$w_{Nw5\,i}$, $P_{Design\_old\,s}$, and $p_{NCw1\,i}$ are defined above, and $p_{NCw2w3w4w5\,i}$ is the probability of non-attrition in any of Waves 2, 3, 4 and 5. For the Wave 3 weight *w3_nc_bp_pweight_s* and the Wave 4 and Wave 5 balanced panel weight, while the final balanced panel weight itself is not scaled, one of the components of the balanced panel weights is the Wave 1 weight, which is scaled in these balanced panel weights' calculations.

6.2.5. Scaling

All the weights for Wave 4 onwards are scaled to the Wave 5 NIDS population total, however, scaling to Wave 5 of NIDS was not undertaken for the original weights for Waves 1, Wave 2 and 3. The probit regressions used to make the non-response and attrition adjustments were also augmented when Wave 4 weights calculations were undertaken, so as to be more consistent across different waves. Scaling and the update to the probit regressions better aligned the Waves 1-4 population sizes and the non-response and attrition adjustments across the waves. Weights for Waves 1, 2 and 3 incorporating both scaling and the updates to the probit regressions have the suffix "_s", as explained in section 6.2.1.

The scaled panel weights are scaled to the NIDS Wave 5 (2017) population total of individuals who were older than 18 during NIDS-CRAM Wave 1, minus the estimated number of emigrants and deaths. This NIDS Wave 5 population total was estimated using NIDS original and top-up members and the Wave 5 post-stratified weights.

The scaling for NIDS-CRAM Wave 1 ensures that all individuals who either successfully completed the interview, or who failed to complete the interview due to moving overseas or dying between NIDS Wave 5 and NIDS-CRAM Wave 1, will have the same population total as the NIDS Wave 5 population total. This counteracts the effect of trimming the NIDS-CRAM Wave 1 weights on the estimated population total.

The scaled panel weights for the original and top-up members in NIDS-CRAM Wave 3 are calculated analogously. The weights reflect the change in design weights due to increased sampling from most strata and are scaled to ensure that the weights of all individuals who either successfully completed the NIDS-CRAM Wave 3 interview, or who failed to complete the Wave 3 interview due to moving overseas or dying between NIDS Wave 5 and NIDS-CRAM Wave 3, will add up to the same population total as the NIDS Wave 5 sample. Note that for original sample members we know whether individuals died or emigrated between NIDS Wave 5 and NIDS-CRAM Wave 1, between Wave 1 and 2 or between Wave 2 and 3; whereas for top-up sample members we only know whether this occurred between NIDS Wave 5 and NIDS-CRAM Wave 3.

The scaled panel weights for Wave 2, Wave 3 (original sample members only) and Wave 4 and 5 (both including and excluding original sample members) are calculated to ensure that all individuals who either successfully completed the Wave t interview (where t indicates 1, 2, 3 or 4), or who failed to complete the interview due to moving overseas or dying between Waves t - 1 and Wave t, will have the same population total as the Wave t - 1 population total (which excludes those who died or moved overseas between Waves t and t - 1). This is achieved by iteratively using the scaled weights for Wave t to estimate the number of people who died between Wave t - 1 and t, and then using this estimate to calculate the population total for Wave t. This population total is then used to calculate the scaled weights for Wave t + 1.

This scaling process is complicated by the fact that some individuals who are classified as not having completed the survey due to having "Moved overseas", successfully complete surveys in subsequent waves. In the calculation of the scaled weights, such individuals are retrospectively declassified as emigrants. This means that any individuals who were classified as emigrants in Wave 4 for example, but who are successfully interviewed in future waves, will affect the calculation of the population totals and hence the scaled waves for all previous waves. The total number of reclassified emigrants tend to be very small, therefore the sampling team anticipate that the effect on the scaled panel weights will be negligible.

6.2.6. Deaths between Wave 1 and Wave 5

Declines in the NIDS-CRAM weighted population between Wave 1 and Wave 5 are due to deaths and emigration. Using the original sample and the scaled weights across the waves, the sampling team estimate that 130,000 individuals died between Wave 1 and Wave 2, 260,000 died between Wave 2 and Wave 3, 245,000 died between Wave 3 and Wave 4 and 55,000 died between Wave 4 and Wave 5. This means that an estimated 690,000 people died between Wave 1 and Wave 5. These estimates are based on relatively small samples and so should be treated with caution - 274 individuals from the sample of 7,073 successfully interviewed in Wave 1 were found to have died by Wave 5.

It is possible to sense check the NIDS-CRAM death estimates against the Medical Research Council (MRC) weekly death data that has been produced since the beginning of the COVID-19 pandemic. The MRC releases weekly death data for individuals aged 1+. NIDS-CRAM surveys only those 18 and older so the numbers are not directly comparable. The NIDS-CRAM surveys for Waves 1-4 were undertaken over 6 weeks and Wave 5 over 4 weeks. This means that using the MRC data to find out exactly how many died between each NIDS-CRAM wave is not straightforward. The MRC data shows that roughly 75,000 people died between NIDS-CRAM Wave 1 and Wave 2, 190,000 died between Wave 2 and Wave 3, 200,000 died between Wave 3 and Wave 4 and 35,000 died between Wave 4 and Wave 5 (there was only a month between the end of Wave 4 and the start of Wave 5). A total of roughly 500,000 people died between NIDS-CRAM Wave 1 and Wave 5.

The NIDS-CRAM estimates are thus around 38% higher, and even slightly larger than 38% when one considers the MRC data includes deaths of individuals aged between 1 and 17. The sampling team's conclusion is that the estimates from NIDS-CRAM are an overestimate but given the small sample size they are not unreasonable.

6.2.7. Household weights[9]

The weights released with NIDS-CRAM are designed to produce estimates for **individuals** – the population covered by those aged 15 years or older in NIDS Wave 5 that could have been interviewed three years later. Nevertheless, there are several variables in NIDS-CRAM which talk about **households**. Because there is potential "double-counting" of households when individual respondents are used, extreme care must be taken in how to interpret those variables. It is always safe to interpret these from the perspective of respondents, but this is probably not easily understood by a general audience.

To avoid some of the double-counting, a set of **household weights** for each of the NIDS-CRAM waves was made available at the same time as the initial NIDS-CRAM Wave 5 release. These weights are included under the technical documents for the NIDS-CRAM data on DataFirst's Open Data Portal, together with the Wittenberg and Branson (2021) paper which discusses them.

As we do not have full household rosters for NIDS-CRAM, these weights adjust double-counting for NIDS Wave 5 households only, i.e., they do not consider any changes in the household distribution since 2017 such as household fission or new household formation. These household weights will be more reliable when used with household variables, but they should still be interpreted with considerable care. Intuitively the weight of the household is "shared" out between those members that could have been interviewed in NIDS-CRAM. There are different ways of doing this and the preferred option is to share it out in proportion to the odds of appearing in the NIDS-CRAM sample. Researchers who want to make use of these weights should consult the technical details in the Branson and Wittenberg (2021) paper to ensure that they understand these properly.

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 8.4, 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

---

9   This section is largely written by Martin Wittenberg.

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
    - o The weights discussed in section 6.2.
- Best estimates of some demographic characteristics of the respondent
    - o These are best estimates of the characteristics considering both NIDS panel and NIDS-CRAM responses.
- Age intervals
    - o Age intervals are calculated from the respondent's best age.
- National lockdown level at the time of interview
- Employment status
    - o 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
    - o These geographic variables are derived from the area in which the respondent reported they were living at the time of interview.
- Interviewer information
    - o 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 version 1.0.0. of the Wave 5 data is as follows:

```
use "Link_File_NIDS-CRAM_Wave5_Anon_1.0.0.dta", clear

merge 1:1 pid using "NIDS-CRAM_Wave5_Anon_1.0.0.dta"

drop _merge

merge 1:1 pid using "derived_NIDS-CRAM_Wave5_Anon_1.0.0.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 12: 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. Important data limitations

8.1. 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).

8.2. 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.

8.3. School attendance

Questions on recent school attendance were excluded from Wave 4. This is because the South African government announced a delay in the start of the 2021 school year shortly before fieldwork began in Wave 4. Schools which were impacted by this announcement would have only opened for the year part-way through the wave, by which stage a large portion of the sample would have already been interviewed.

8.4. 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:
    - o   In-person interviews in NIDS and CATI in NIDS-CRAM.

- Questionnaire type differences:
  - o 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.
  - o 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.

8.4.1. Derived and imputed variables

8.4.1.1. 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.

8.4.1.2. 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.

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

8.4.2. 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.

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

8.4.4. 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.

8.4.5. 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.

8.4.6. 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.

9. 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, if necessary.
- 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.

9.1. Changes across public release versions of NIDS-CRAM

9.1.1. Wave 1: V1.1.0 to V2.0.0

1.

Dataset	NIDS-CRAM
Variable	*w1_nc_emreturn*
Change	98 recoded to "-9. Don't know" and 99 recoded to "-8. Refused".
Reason for change	Recoded the non-response values to standard NIDS-CRAM non-response values.
Number of observations impacted	248 "Don't Know" & 104 "Refused" cases: Total 352.

2.

Dataset	Derived
Variable	*w1_nc_pweight_s*
Change	Added to derived data file.
Reason for change59,831	Weights were scaled to the NIDS Wave 5 population totals and the weights algorithms were updated.
Number of observations impacted	All.

3.

Dataset	All datasets
Variable	N/A
Change	Removed the "W1" from the file name of all files as it was "Wave1_W1".
Reason for change	Redundant wave name in file name.
Number of observations impacted	N/A

4.

Do file	Program 1 - NIDS-CRAM - Deflators_NIDS_W5
Change	Changed name from "Program 1a - Deflators_NIDS_W5".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

5.

Do file	Program 1a - NIDS-CRAM - Deflators_W1
Change	Changed name from "Program 1b - Deflators_NIDS-CRAM_W1".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

6.

Do file	Program 2a - NIDS-CRAM - Employment_Status_W1
Change	Changed name from "W1_NIDS-CRAM_Employment_Status".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

7.

Do file	Program 3 - NIDS-CRAM - Merging NIDS and NIDS-CRAM
Change	Changed name from "Program 1 - Merging NIDS and NIDS-CRAM". Removed deflators incorporation.
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves. Deflators inclusion was not implied by the file name or needed for the purpose of this do file.

8.

Do files	Program 4 - NIDS-CRAM - Merging files for a given wave into a cross-section Program 5 - NIDS-CRAM - Merging all the NIDS-CRAM data into a panel
Change	Added.
Reason for change	Provision of same types of do files in each wave.

9.1.2. Wave 1: V2.0.0 to V3.0.0

1.

Dataset	NIDS-CRAM
Variable	All
Change	Minor data cleaning.
Reason for change	Minor data quality errors being corrected.
Number of observations impacted	All.

9.1.3. Wave 2: V1.1.0 to V2.0.0

1.

Dataset	NIDS-CRAM
Variable/s	w2_nc_ems_june (and some variables thereafter in Section C of the questionnaire).
Change	Some cases needed to change from "2. No - I'm employed" to "No, I did nothing".
Reason for change	Due to a Zulu translation error.
Number of observations impacted	10
Notes	The result of following the skip patterns was that the next questions in the skip sequence needed to be changed to "-3. Missing" or system missing, where appropriate.

2.

Dataset	Derived
Variable	*w2_nc_empl_stat*
Change	10 observations changed from "3. Employed" to "-8. Refused".
Reason for change	This is a derived variable and the underlying variables changed due to a translation error in Zulu.
Number of observations impacted	10

3.

Dataset	Derived
Variable	*w2_nc_pweight_s*
Change	Added to derived data file.
Reason for change	Weights were scaled to the NIDS Wave 5 population totals and the weights algorithms were updated.
Number of observations impacted	All.

4.

Do file	Program 1b - NIDS-CRAM - Deflators_W2
Change	Changed name from "Program 1c - Deflators_NIDS-CRAM_W2".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

5.

Do file	Program 2b - NIDS-CRAM - Employment_Status_W2
Change	Changed name from "W2_NIDS-CRAM_Employment_Status".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

6.

Do file	Program 4 - NIDS-CRAM - Merging files for a given wave into a cross-section
Change	Changed name from "Program 2 - Merging files for a give wave into a cross-section".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

7.

Do files	Program 5 - NIDS-CRAM - Merging all the NIDS-CRAM data into a panel
Change	Changed name from "Program 3 - Merging all the NIDS-CRAM data into a panel". Changed hardcoded latest wave to dynamic latest wave.
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves. The dynamic latest wave obviates the need to change the latest wave number.

8.

Do files	Program 1 - NIDS-CRAM - Deflators_NIDS_W5 Program 3 - NIDS-CRAM - Merging NIDS and NIDS-CRAM
Change	Added.
Reason for change	Provision of same types of do files in each wave.

9.1.4. Wave 2: V2.0.0 to V3.0.0

1.

Dataset	NIDS-CRAM
Variable	All.
Change	Minor data cleaning.
Reason for change	Minor data quality errors being corrected.
Number of observations impacted	7051

9.1.5. Wave 3: V1.0.0 to V2.0.0

1.

Dataset	NIDS-CRAM
Variable	*w3_nc_ems_oct* (and some variables thereafter in Section C of the questionnaire).
Change	Some cases needed to change from "2. No - I'm employed" to "No, I did nothing".
Reason for change	Due to a Zulu translation error.
Number of observations impacted	20
Notes	The result of following the skip patterns was that the next questions in the skip sequence needed to be changed to "-3. Missing" or system missing, where appropriate.

2.

Dataset	Derived
Variable	*w3_nc_empl_stat*
Change	20 observations changed from "3. Employed" to -8. "Refused".
Reason for change	This is a derived variable and the underlying variables changed due to a translation error in Zulu.
Number of observations impacted	20

3.

Dataset	NIDS-CRAM
Variable	All variables after section A.
Change	Data set to system missing.
Reason for change	5 non-response respondents had some data filled in for the first few questions. In error, these were not previously set to missing, as is required.
Number of observations impacted	5

4.

Dataset	Derived
Variable	*w3_nc_pweight_s*
Change	Added to derived data file.
Reason for change	Weights were scaled to the NIDS Wave 5 population totals and the weights algorithms were updated.
Number of observations impacted	All.

5.

Dataset	Derived
Variable	*w3_nc_pweight_extu_s*
Change	Added to derived data file.
Reason for change	Weights were scaled to the NIDS Wave 5 population totals and the weights algorithms were updated.
Number of observations impacted	All.

6.

Dataset	Derived
Variable	*w3_nc_bp_pweight_s*
Change	Added to derived data file.
Reason for change	Weights were scaled to the NIDS Wave 5 population totals and the weights algorithms were updated. One component of this weight is the Wave 1 weight, which is scaled in this version of the balanced panel weight.
Number of observations impacted	All.

7.

Do file	Program 1c - NIDS-CRAM - Deflators_W3
Change	Changed name from "Deflators_NIDS-CRAM_W3".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

8.

Do file	Program 2c - NIDS-CRAM - Employment_Status_W3
Change	Changed name from "Deflators_NIDS-CRAM_W3".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

9.

Do file	Program 4 - NIDS-CRAM - Merging files for a given wave into a cross-section
Change	Changed name from "Program 2 - Merging files for a give wave into a cross-section".
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves.

10.

Do files	Program 5 - NIDS-CRAM - Merging all the NIDS-CRAM data into a panel
Change	Changed name from "Program 3 - Merging all the NIDS-CRAM data into a panel". Changed hardcoded latest wave to dynamic latest wave.
Reason for change	Differentiating program numbers from NIDS and improving numbering across NIDS-CRAM waves. The dynamic latest wave obviates the need to change the latest wave number.

11.

Do files	Program 1 - NIDS-CRAM - Deflators_NIDS_W5 Program 3 - NIDS-CRAM - Merging NIDS and NIDS-CRAM
Change	Added.
Reason for change	Provision of same types of do files in each wave.

9.1.6. Wave 3: V2.0.0 to V3.0.0

1.

Dataset	Derived
Variable	*w3_nc_pweight_s*
Change	Weights updated.
Reason for change	Weights updated as a result of outcomes in Wave 5.
Number of observations impacted	6130

2.

Dataset	Derived
Variable	*w3_nc_pweight_extu_s*
Change	Weights updated.
Reason for change	Weights updated as a result of outcomes in Wave 5.
Number of observations impacted	6130

3.

Dataset	Derived
Variable	*w3_nc_bp_pweight_s*
Change	Weights updated.
Reason for change	Weights updated as a result of outcomes in Wave 5.
Number of observations impacted	4508

4.

Dataset	NIDS-CRAM
Variable	All
Change	Minor data cleaning.
Reason for change	Minor data quality errors being corrected.
Number of observations impacted	8098

9.1.7. Wave 4: V1.0.0 to V2.0.0

1.

Dataset	Derived
Variable	*w4_nc_pweight_s*
Change	Weights updated.
Reason for change	Weights updated as a result of outcomes in Wave 5.
Number of observations impacted	5629

2.

Dataset	derived
Variable	*w4_nc_pweight_extu_s*
Change	Weights updated
Reason for change	Weights updated as a result of outcomes in Wave 5.
Number of observations impacted	5629

3.

Dataset	derived
Variable	*w4_nc_bp_pweight_s*
Change	Weights updated
Reason for change	Weights updated as a result of outcomes in Wave 5.
Number of observations impacted	**3837**

4.

Dataset	NIDS-CRAM
Variable	All
Change	Minor data cleaning.
Reason for change	Minor data quality errors being corrected.
Number of observations impacted	5629

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.

Wittenberg, M. and Branson, N. (2021). Creating Household Weights for NIDS-CRAM. Forthcoming NIDS-CRAM working paper.