D

WAVE 2

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

NIDS-CRAM Wave 1 Data Quality

Cally Ardington - University of Cape Town

6 September 2020

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

CRAM CORONAVIRUS RAPID MOBILE SURVEY 2020

NIDS-CRAM Wave 1 Data Quality

Cally Ardington

6 September 2020

Data Access

The NIDS-CRAM data is publicly available for download and use from the DataFirst portal

www.datafirst.uct.ac.za

1. Introduction

This technical report assesses the overall quality of data collected in the first wave of NIDS-CRAM. Several quality dimensions are investigated including the importance of item non-response, measurement error and the degree to which the data on key variables adequately represent the national situation. The focus on these data-quality dimensions is motivated by several considerations. Most obvious is the usefulness of the data in informing policy. To that end this paper with its focus on item level data quality, complements an earlier technical paper that considered the issue of representivity at the level of the sample frame and unit non-response. Surveys around individual and household wellbeing are typically conducted face-to-face. The lockdown in response to Covid-19 forced researchers into new modes of data collection and a consideration of NIDS-CRAM data quality is informative in assessing the feasibility of telephone surveys for policy relevant social science research.

Several NIDS-CRAM Wave 1 discussion papers mention issues of data quality but within thematic silos and generally as technical details. This report draws on these papers but here these issues are foregrounded and consolidated into one easily accessible resource.

The paper begins with an overview of item non-response in Section 2. With the important exceptions of household income and earnings, item level non-response does not pose a significant problem. Section 3 considers the accuracy and descriptive representativeness of key variables by comparing their characteristics with those from appropriate benchmark samples. Where results do not align with other surveys, including previous waves of NIDS, the paper investigates whether this can be attributed to differences in questionnaire design, data collection methods, non-response or possibly related to the impact of Covid-19. The following section returns to focus in more detail on the household income data. The extensive information gathered prior to NIDS-CRAM, together with non-missing Wave 1 variables is used to investigate potential bias due to non-response and measurement error. Section 5 concludes.

2. Item non-response

Item non-response occurs when a respondent refuses to answer the question or does not know the answer. In some instances, a “don’t know” response is a valid outcome of interest (e.g. for questions about knowledge of Covid-19 symptoms). In others, it may be a soft refusal[1] or indicative of a poorly worded question. It may also reflect the uncertain nature of the item (e.g. income from survivalist enterprises) or the times (e.g. employment status under lockdown).

In addition to the impact of the unprecedented shock of Covid-19, there are other possible reasons, such as the mode of the survey or the selected respondent, why non-response may be higher in NIDS-CRAM than in other South African surveys. Telephone surveys are substantially shorter and offer less chance for the interviewer to establish a rapport with respondents than face-to-face interviewing. There may also be greater privacy issues with phone surveys, particularly during the lockdown period. Typically, household level questions are asked of the household member who is most knowledgeable around income, expenditure etc. NIDS-CRAM is a panel of individuals who will vary in their suitability to answer questions about their households. This is likely exacerbated by the movement of individuals between households in anticipation of the lockdown.

The extent to which non-response may bias estimates depends on the non-response mechanism but, if prevalence is low then even non-random non-response will not pose a problem. This paper therefore begins by methodically examining the item non-response rate for each variable in NIDS-CRAM Wave 1. Of the 139 variables, 36 had non-response rates in excess of five percent. These variables are shown in Appendix *Table A1*. However, many of these variables follow skip patterns and are asked of very few respondents. It is more appropriate and informative to examine item

1 A soft refusal occurs when a respondent is knowledgeable about the item but gives a don’t know response as a softer way of refusing to answer the question than an explicit refusal.

non-response for the derived variables for which these variables serve as building blocks (e.g. employment and earnings). *Table 1* presents item non-response rates for several key derived variables and a few selected questions from *Table A1* that were posed to the majority of respondents.

Table 1. Item non-response in NIDS-CRAM Wave 1

Description	Variable(s)	% missing	N
Total household income after tax in April	w1_nc_hhinc	37.7%	7073
Do you think you are likely to get the Coronavirus?	w1_nc_cvrsk	12.9%	7073
What are some of the symptoms of Coronavirus?	w1_nc_cvsypt1	10.0%	7073
Last 7 days, did children in household:			
access educational content online?	w1_nc_edccon	8.9%	5775
watch educational TV?	w1_nc_edctv	6.7%	5775
listen to educational radio?	w1_nc_edcrad	6.7%	5775
use their school books?	w1_nc_edcbook	5.6%	5775
Employment:			
Employed February	w1_nc_em_feb, w1_nc_emany_feb, w1_nc_ems_feb	1.3%	7073
Employed April	w1_nc_em_apr, w1_nc_emany_apr, w1_nc_ems_apr	2.0%	7073
Employed April (including do you have a job to return to)	w1_nc_em_apr, w1_nc_emany_apr, w1_nc_ems_apr, w1_nc_emreturn	5.3%	7073
Earnings:			
Earnings February	w1_nc_eminc_feb, w1_nc_eminc_feb_brac	6.0%	3408
Earnings April	w1_empay_freq, w1_nc_eminc_apr, w1_nc_eminc_apr_brac, w1_nc_emwkinc_apr, w1_nc_emwkinc_apr_brac, w1_nc_em2wksinc_apr, w1_nc_em2wksinc_apr_brac, w1_nc_emdayinc, w6_emdays_apr, w1_nc_emsinc_apr, w1_emsinc_apr_brac	12.4%	2752
Social grants:			
Number of CSG in household	w1_nc_hhincchld, w1_nc_incchld	2.5%	7073
Number of OAP in household	w1_nc_hhincgovpen, w1_nc_incgovpen	0.9%	7073
Any CSG in household	w1_nc_hhincchld, w1_nc_incchld	1.3%	7073
Any OAP in household	w1_nc_hhincgovpen, w1_nc_incgovpen	0.8%	7073
Individual grant receipt	w1_nc_incgov, w1_nc_incgovtyp1, w1_nc_incgovtyp2, w1_nc_incgovtyp3	0.6%	7073

The highest non-response rate is for the question around after tax household income tax in April at around 38 percent. While this is unsurprising given the typically high rates of non-response for questions around income, it does present a substantive challenge for researchers and policymakers using NIDS-CRAM. Non-response for the household income variable will be revisited in some detail, together with other data quality issues for this variable in section 4 below.

From the health section of the survey, there were two questions with substantial non-response but in these instances "don't know" would be considered as a valid response rather than missing data. One in ten respondents did not know any of the symptoms of Coronavirus and around 13 percent did not have a sense of how likely they were to get the Coronavirus.

Questions about children in the household engaging with educational content on TV, radio and online had non-response rates ranging from six to nine percent. Non-response rates are higher amongst males than females suggesting that non-response here is related to respondents' familiarity with the educational activities of children in the household.

Employment status and earnings are derived from a range of variables and non-response can be due to missing data on any of these input variables. Non-response is not an issue for deriving employment status unless one includes the question about having a job to return to after the lockdown around which respondents indicate substantial uncertainty.

For the earnings questions, respondents who could not or would not indicate a specific amount, were given the opportunity to report their earnings in brackets. Around 16 and 11 percent of individuals provided a bracketed response for February and April earnings respectively. Despite this, the derived earnings are missing for six percent of those employed in February and for 12 percent of those employed in April. These non-response rates for earnings are not that dissimilar to NIDS Wave 5 where nine percent of employed individuals aged 18 or older have missing earnings data. There is a very low rate of missing data for earnings (1.8%) in the 2018 General Household Survey (GHS). However, 41 percent of GHS respondents provided a bracketed response. While the substantially higher non-response rate in April is unsurprising given the uncertainty surrounding lockdown, it may also be driven by the larger number of input variables. There is likely a trade-off between the non-response rate and the accuracy of non-missing responses depending on the number of earnings related questions asked.

The role of social grants in ameliorating the impact of lockdown has been a key focus of NIDS-CRAM researchers and non-response rates on key grant variables are presented in the final rows of *Table 1*. Here non-response does not present a challenge.

In general, response rates for this telephone survey are fairly similar to those from the face-to-face interviews in previous waves of NIDS. For example, around 30 percent of the 868 variables[2] from the NIDS Wave 5 adult interview have non-response rates of five percent or higher in contrast to 26 percent of the NIDS-CRAM variables. With the important exception of household income, item level non-response does not pose a significant problem for NIDS-CRAM. The paper returns to examine this variable in detail in section 4 below.

3. Comparability with other surveys

The degree to which NIDS-CRAM adequately represents the national situation is an important consideration for the usefulness of NIDS-CRAM in informing policy responses and debates. Earlier technical papers (Kerr et al. 2020a, 2020b) considered the issue of representivity of NIDS-CRAM at the level of the sample frame and unit non-response. Here the focus is on representivity at the item level. The unprecedented shock of Covid-19 presents a challenge for this exercise as the impact on many key outcomes is unknown. Indeed, estimating that impact is the very rationale for NIDS-CRAM. Nevertheless, comparisons on certain key variables against appropriate benchmark surveys can alert us to where there may be data quality issues and provide suggestive evidence on the extent to which levels and trends observed in NIDS-CRAM are a relevant reflection of dynamics in the South African population.

This section begins by comparing characteristics of NIDS-CRAM correspondents against earlier

2 This includes only variables directly related to questions asked of the respondent.

waves of NIDS and against a recent Stats SA dataset[3]. *Table 2* presents means for a range of household level variables for NIDS-CRAM Wave 1 and the comparison samples. The first comparison is actually within the NIDS-CRAM sample with data from NIDS Wave 5 giving an indication of how NIDS-CRAM respondent characteristics differ between 2017 and 2020. These estimates use NIDS Wave 5 data with NIDS-CRAM Wave 1 weights and are indicated by NIDS Wave 5 (CRAM weights) in the tables and figures. The two comparison samples are the full NIDS Wave 5 sample and GHS 2018. These samples are restricted to individuals aged 18 and over. Weighted estimates are provided for all samples. NIDS-CRAM is a panel of individuals rather than households. The unit of analysis for these comparisons is therefore the individual for both individual and household level variables – i.e. the electricity variable represents the percentage of individuals older than 18 years of age living in a household with electricity.

Table 2. Household characteristics for individuals age 18 and older – NIDS-CRAM and benchmark samples

	NIDS-CRAM Wave 1	NIDS Wave 5 (CRAM weights)	NIDS Wave 5	GHS 2018
Electricity	0.95	0.90	0.90	0.96
Piped water	0.83	0.78	0.78	0.76
Dwelling type:				
House or flat	0.80	0.81	0.81	0.85
Traditional (e.g. mud)	0.08	0.07	0.07	0.06
Informal (e.g. shack)	0.11	0.11	0.12	0.09
Other	0.01	0.00	0.00	0.01
Household size	4.94	4.29	4.24	4.70
Number of children 0 to 6	0.72	0.64	0.65	0.67
Number of children 7 to 17	0.99	0.97	0.89	0.91
Number of adults 60 plus	0.52	0.35	0.34	0.38
Household income sources				
Employment	0.53	0.64	0.64	0.77
Grants	0.42	0.52	0.52	0.57
Family & friends / remittances	0.11			0.13
Other	0.03			0.08
None	0.06			0.00
At least one CSG	0.53	0.50	0.49	0.45
At least one OAP	0.32	0.24	0.23	0.22
Urban	0.83	0.66	0.67	0.69

Comparing the first two columns in *Table 2* shows that there has been a slight increase in the percentage of individuals living in households with electricity and piped water as would be expected over the three-year period between the studies. Household size and the number of adults aged 60

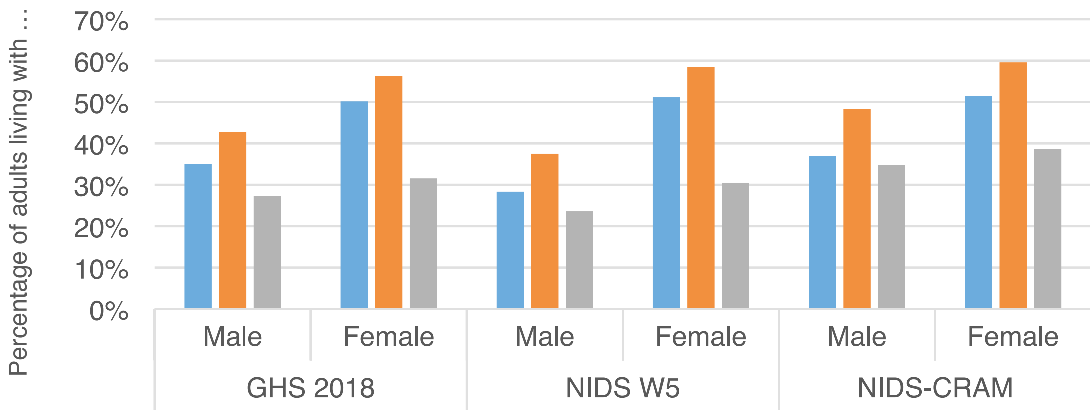
3 This section draws on Wills et al. 2020, Jain et al. 2020, Ranchhod and Daniels 2020 and van der Berg et al. 2020 both consolidating and expanding on the data quality analyses of those papers.

and older tends to be higher in NIDS-CRAM than in all other samples, but this could plausibly be a result of household re-arrangements around lockdown (Posel and Casale 2020). Consistent with higher average number of older adults, the percentage of households receiving at least one state old age pension is higher in NIDS-CRAM. The percentage of individuals living in households that receive at least one child support grant is very similar across the samples. NIDS-CRAM respondents are substantially less likely than individuals in the GHS to report earnings from employment as a source of household income. It is difficult to know what to make of this divergence between the samples. These differences could be driven by issues such as survey implementation (e.g. knowledgeable household respondent versus randomly sampled individual), changing living arrangements in response to lockdown and the economic impact of lockdown.

Around 83 percent of individuals in NIDS-CRAM are living in an urban area, substantially higher than the 66 to 69 percent of the other samples. For the NIDS-CRAM sample, 63 percent of individuals living in a rural area in 2017 are coded as living in an urban area. This degree of urbanisation over three years seems implausible and cannot be explained by movements in anticipation of the lockdown[4]. In previous waves of NIDS, the interviewer recorded the GPS co-ordinates of the household where the interview took place. In NIDS-CRAM, the geographic variables are derived from the area where the respondent reported living at the time of interview (Ingle et al. 2020). Individuals who report living in the area of a town or city are recorded as urban, even though they may be living in rural surrounding areas. This method of classification has introduced a degree of measurement error and researchers should be cautious in using the NIDS-CRAM geographic location type variable to compare urban and rural respondents.

Before turning to comparability of individual characteristics, differences in household composition are examined through a gender lens. *Figure 1* shows the percentage of individuals aged 18 and older living with small children, school aged children and older adults separately for men and women for the GHS, NIDS Wave 5 and NIDS-CRAM. The gendered patterns in living arrangements are clear and remarkably similar across surveys, with more women living with small and school aged children than men. The data suggest that both men and women are more likely to be living with older adults during lockdown than in 2017/2018.

Figure 1. Percentage of individuals living with children and older adults



	GHS 2018 Male	GHS 2018 Female	NIDS W5 Male	NIDS W5 Female	NIDS-CRAM Male	NIDS-CRAM Female
Children 0-6	35%	50%	28%	51%	37%	51%
Children 7-17	43%	56%	38%	58%	48%	60%
Adults 60+	27%	32%	24%	30%	35%	39%

4 Around 10 percent of all respondents report moving in anticipation of the lockdown. This figure is 9 percent for individuals who changed status from rural to urban between NIDS Wave 5 and NID-CRAM Wave 1. Around 78 percent of individuals who are coded as having moved from a rural to an urban area are living in the same district council in 2017 and 2020.

In general, results in *Table 2* do not raise any red flags with respect to comparability across surveys. Turning to individual level characteristics in *Table 3*, the NIDS-CRAM sample appears similar to the other samples on basic demographic variables of age[5], sex and race and on the receipt of the state old age pension and disability grants. Medical aid coverage is slightly higher in NIDS-CRAM but largely in line with the other samples. A shift from incomplete secondary to completing matric is expected for NIDS-CRAM respondents between 2017 and 2020 (comparing columns one and two) but the percentage of matriculants in NIDS-CRAM does appear high in comparison to the full NIDS sample and the GHS. The employment rate in NIDS-CRAM[6] and NIDS Wave 5 is higher than in the GHS while receipt of the child support grant is implausibly low. Each of these issues is examined in turn.

Table 3. Individual characteristics for individuals age 18 and older – NIDS-CRAM and benchmark samples

	NIDS-CRAM Wave 1	NIDS Wave 5 (CRAM weights)	NIDS Wave 5 (NIDS weights)	GHS 2018
Age	40.4	37.1	38.9	39.2
Female	0.53	0.53	0.52	0.52
African	0.78	0.80	0.81	0.79
Coloured	0.10	0.09	0.09	0.09
Asian/Indian	0.02	0.02	0.02	0.03
White	0.10	0.09	0.08	0.10
Employed (18 – 59 year olds)	0.58	0.55	0.55	0.53
Education level				
None	0.02	0.04	0.05	0.04
Up to primary	0.13	0.12	0.13	0.13
Up to secondary	0.35	0.45	0.44	0.39
Matric	0.51	0.40	0.38	0.44
Medical aid	0.22	0.17	0.16	0.18
CSG recipient	0.06	0.20	0.21	
OAP recipient	0.09	0.09	0.09	0.09
Disability grant recipient	0.02	0.02	0.03	0.02

5 The NIDS-CRAM respondents were three years younger at time of NIDS Wave 5. This is reflected in the lower average age in the second column.

6 The employment rate for NIDS-CRAM is for employment in February 2020.

Employment

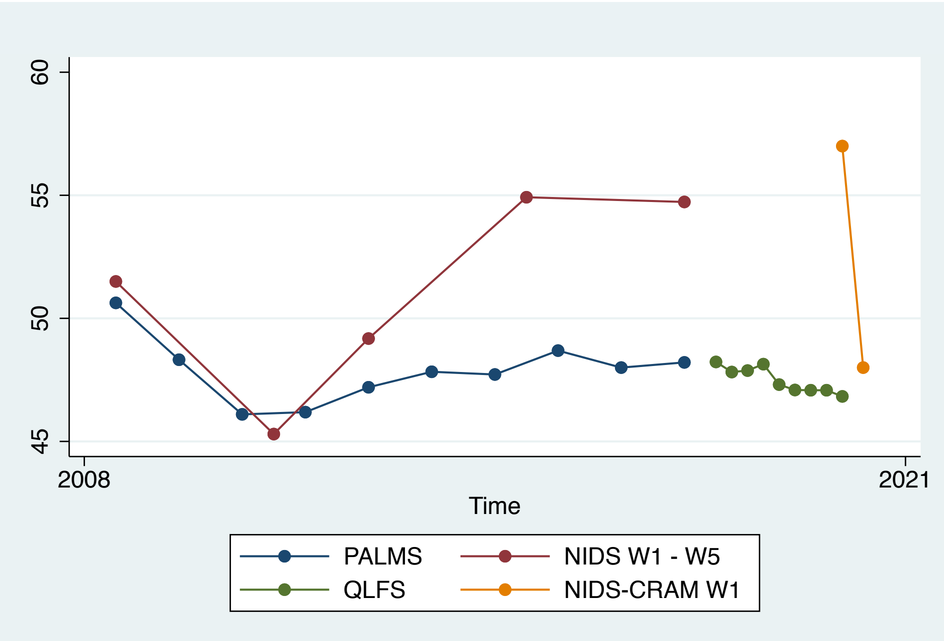
Table 4 summarises how the differences in employment rates translate to differences in employment numbers. Estimates of the total number of employed individuals age 18 to 59, together with 95 percent confidence intervals are shown for NIDS-CRAM Wave 1 and the GHS 2018. NIDS-CRAM estimates suggest that around 450,000 more individuals were employed in February 2020 than in the last quarter of 2018. However, the confidence interval for NIDS-CRAM in fairly wide and overlaps with that from the GHS 2018. Interestingly, the GHS 2018 produces a higher estimate for the total number of employed men than NIDS-CRAM.

Table 4. Total number of employed individuals aged 18 to 59 years of age

	NIDS-CRAM Wave 1	GHS 2018
All	17,482,190	17,023,470
	(15684640;	(16636160;
	19279730)	17410780)
Males	9323776	9649666
	(8292029;	(9412737;
	10355520)	9886595)
Females	8133849	7373801
	(7219166;	(7167990;
	9048532)	7579612)

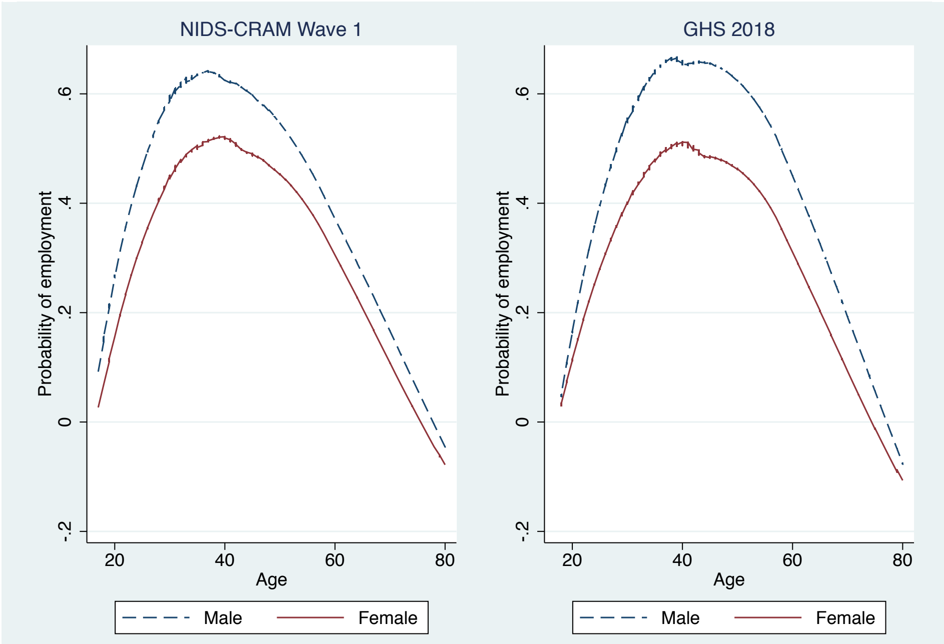
The issue of a high pre-lockdown employment rate in NIDS-CRAM has been addressed in detail by Ranchhod and Daniels (2020) and Kerr et al. (2020b). *Figure 2*, reproduced from Ranchhod and Daniels, clearly shows that the NIDS employment rate was initially aligned with estimates from Stats SA datasets and then diverged over time due to non-random attrition. Given that NIDS-CRAM was sampled from NIDS Wave 5, any bias remaining after adjustments for attrition carries through.

Figure 2. Employment rate for adults aged 18 to 59



While such biases can be important for precise estimates of job loss, trends identified in NIDS-CRAM are indicative of current labour market dynamics in South Africa (Kerr et al. 2020b). Indeed, the value of NIDS-CRAM should not be restricted to counting the exact number of unemployed or hungry South Africans. The detailed longitudinal data offers an opportunity to understand dynamics and examine the interplay of factors that mitigate and exacerbate vulnerabilities and inequities. With this in mind, this paper examines not only levels but also relationships between variables in the assessment of data quality. *Figure 3* below presents non-parametric estimates of the probability of employment by age and sex for NIDS-CRAM and the GHS. Despite, substantial differences in the level of employment, the employment-age gradients are remarkably similar across the two studies.

Figure 3. Probability of employment by age and sex – NIDS-CRAM Wave 1 and GHS 2018



Social grants

Compared to NIDS Wave 5, implausibly few NIDS-CRAM respondents report child support grant receipt. However, this is not a sample representivity issue but rather driven by a particular (mis) understanding of the question. Unlike other government grants, there is a distinction between the recipient (caregiver) and the beneficiary (child) of a child support grant. NIDS-CRAM first asked respondents "How many child support grants does this household receive?" and then a few questions later "Which government grant (or grants) do you receive?" The ordering and wording of these questions appears to have resulted in under-reporting individual receipt of the child support grant. In contrast, the question on the number of child support grants that the household receives produces estimates that are much more in line with expectations. *Figures 4 to 7* shed further light on this issue with estimates of the total number of grant recipients from each sample together with the South African Social Security Agency figures on the number of grant beneficiaries in April 2020 (SASSA 2020). In each figure the SASSA figure is shown by the orange line. Point estimates are indicated by the black square and the lines indicate the 95 percent confidence interval. The confidence intervals for the number of state old age pension recipients are overlapping for all four samples (*Figure 4*). All confidence intervals are below the SASSA figure but are within the ballpark.

Figure 4. State Old Age Pension Recipients

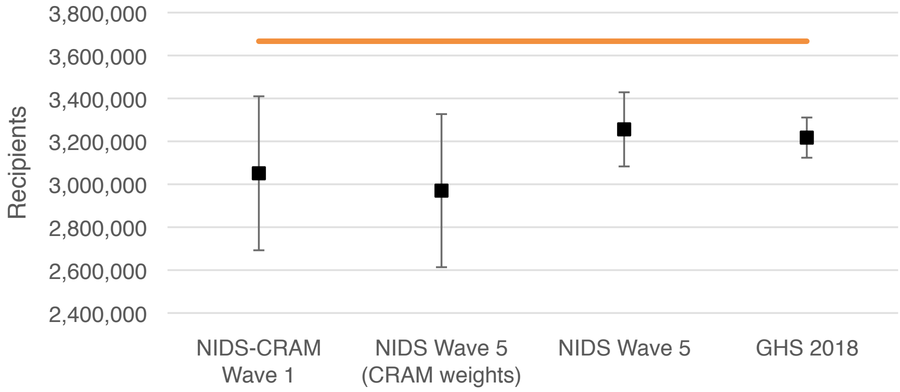
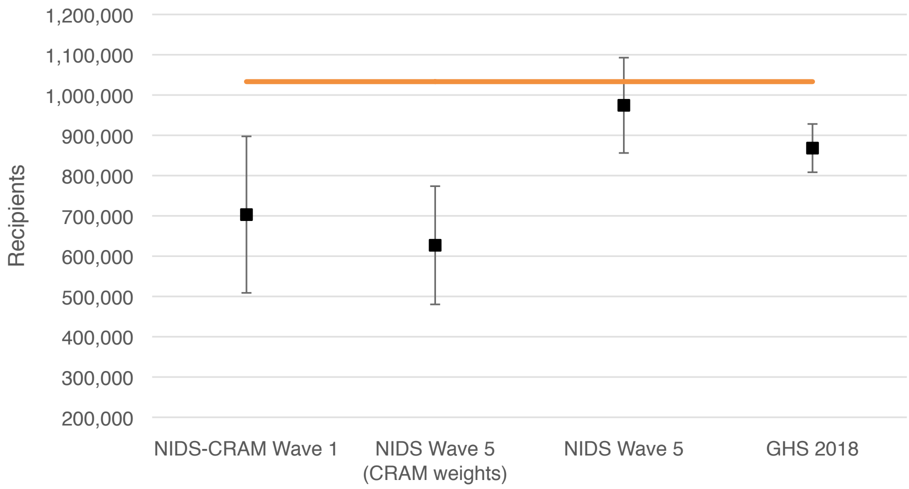


Figure 5 shows estimates of the number of disability grant recipients. The estimates for NIDS-CRAM sample using NIDS-CRAM Wave 1 or NIDS Wave 5 data are lower than those for the full NIDS Wave 5 sample and the GHS, which are more in line with the SASSA figure. Nevertheless, the confidence interval for NIDS-CRAM overlaps with that for NIDS and the GHS.

Figure 5. Disability grant recipients



As expected, given the low proportion of respondents reporting child support grant receipt in *Table 3*, the estimate of the number of child support recipients in NIDS-CRAM is much too low (*Figure 6*). Estimates from NIDS Wave 5 are very much in line with SASSA figures[7]. This confirms that the under-estimate is not due to sample bias but rather to the way in which individuals in NIDS-CRAM responded to the question. *Figure 7* uses data on children under 18 years of age from NIDS Wave 5 and the GHS to estimate the number of child beneficiaries of the child support grant. Estimates are similar in both samples and somewhat higher the SASSA figure. This provides some assurance that the NIDS data is broadly representative with respect to child support grant caregivers and beneficiaries.

7 The GHS records child support grants under the beneficiary (child) and not the recipient (caregiver).

Figure 6. Child support grant caregivers

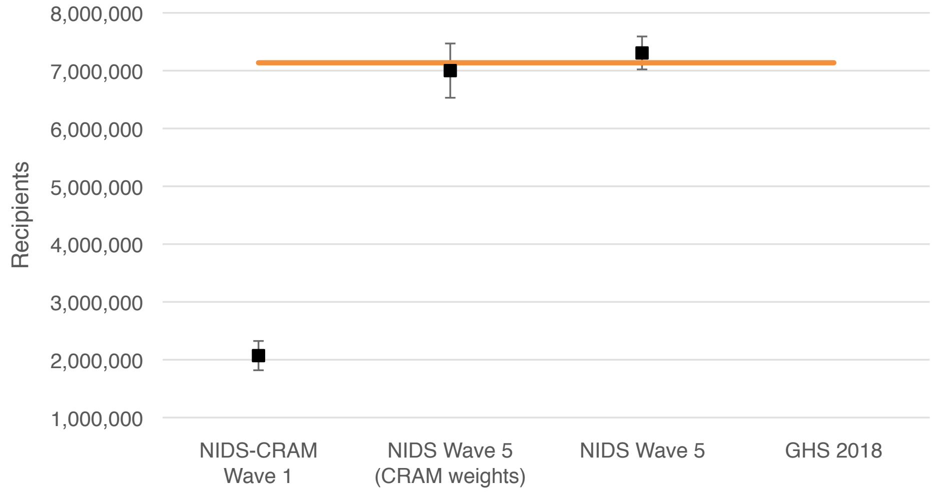


Figure 7. Child support grant child beneficiaries

NIDS-CRAM also included questions on the number of child support grant and old age pension beneficiaries in the household. Four percent of individuals report a greater number of state old age pensions in the household than the number of adults aged 60 and older living in the household. Eight percent of individuals report a greater number of child support grants than the number of children currently living in the household. It is possible that caregivers in the household are receiving the child support grant for children living elsewhere but these discrepancies suggest some degree of measurement error in one or both of these variables. *Table 5* presents estimates, together with 95 percent confidence intervals, of the total number of child support grant and old age pension beneficiaries living with the adults that make-up the four samples. Beneficiaries will be counted for every adult individual with whom they co-reside. The totals for NIDS-CRAM child support grants are in line with both earlier waves of NIDS and Stats SA's GHS. The estimate of the number of state old age pensions in the household is substantially higher for NIDS-CRAM than for the other samples[8]. This accords with the NIDS-CRAM respondents being more likely to co-reside with adults aged 60 and older (*Figure 1* and *Table 2*).

8 Replacing the number of state old age pensions with the number of adults aged 60 and older in the cases where the former is greater, reduces the estimate to 12,648,820.

Table 5. Total child support grant and old age pension beneficiaries in the household

	NIDS-CRAM Wave 1	NIDS Wave 5 (CRAM weights)	NIDS Wave 5 (NIDS weights)	GHS 2018
Total CSGs in household	41,735,900 (39,729,490; 43,742,360)	39,821,760 (37,989,830; 41,653,690)	38,256,700 (37,199,700; 39,313,700)	42,047,570 (41,387,950; 42,707,180)
Total OAPs in household	14,253,300 (13,344,250; 15,162,290)	9,903,173 (9,221,663; 10,584,68)	9,817,891 (9,485,550; 10,150,230)	10,408,120 (10,208,820; 10,607,410)

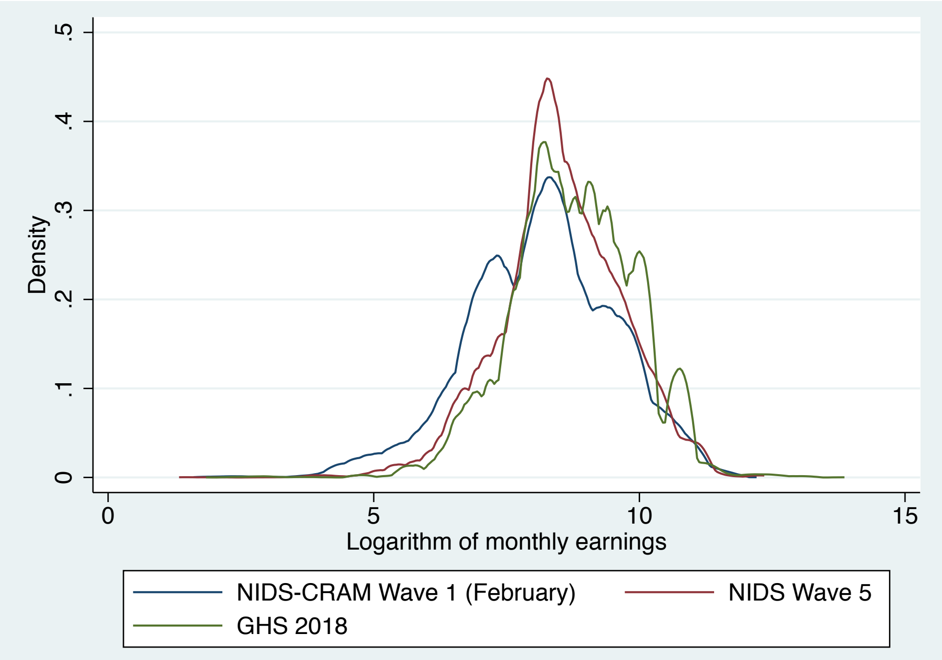
In summary, with the exception of child support grant caregivers, estimates of individual grant receipt are broadly similar across samples and provide reasonable estimates relative to SASSA figures.

Earnings

Comparisons of earnings data are limited by the absence of any priors on the distribution of wages during lockdown and the focus is therefore restricted to the earnings distribution in February 2020. *Figure 8* plots the weighted distributions of the logarithm of monthly earnings for individuals aged 18 and older. All earnings are inflated to February 2020. In the 16 percent of cases where individuals provided a bracket response for earnings, individuals were assigned the median earnings of individuals who had given a specific value in the range of the bracket. The left peak for the February 2020 earnings distribution corresponds to the imputed value for the bottom earnings bracket. This bottom bracket ranged from R1 to R3000 and is significantly wider for NIDS-CRAM than NIDS Wave 5 and the GHS, which have five to six brackets up to R3000. Within the range R1 to R3000, bracketed responses in NIDS Wave 5 and the GHS were concentrated at the higher end. The wide bottom bracket in NIDS-CRAM may lead to downward biased estimates in the range below R3000.

Comparing across samples, the distribution of earnings for the GHS is to the right of that for NIDS Wave 5, which in turn is to the right of that for NIDS-CRAM February 2020. Nevertheless, the distributions are not too dissimilar – there is considerable overlap between distributions and modal values are closely located. This provides some confidence that the earnings data from February are plausible given the distribution of earnings in the benchmark samples.

Figure 8. Distribution of monthly earnings – GHS 2018, NIDS wave 5 and NIDS-CRAM Wave 1



As a further validation of NIDS-CRAM earnings data, covariates from earnings regressions are compared across samples. Here the concern is less with the level of earnings and more on the association between earnings and various individual characteristics. *Table 6* presents results from OLS regressions of the logarithm of earnings on age, sex, race and education for NIDS-CRAM (February and April earnings), NIDS Wave 5 and the GHS. Within each sample, the sign and relative magnitude of the coefficients align with previous research on earnings in South Africa. The coefficients are remarkably similar across samples.

Table 6. Comparison of regression results across datasets – logarithm of monthly earned income

Dependent variable: Logarithm of monthly earnings	NIDS-CRAM February	NIDS-CRAM April	NIDS Wave 5	GHS 2018
Age	0.0175***	0.0145***	0.0124***	0.00835***
	(0.00205)	(0.00229)	(0.000848)	(0.000646)
Male	0.479***	0.505***	0.484***	0.312***
	(0.0414)	(0.0460)	(0.0185)	(0.0141)
Coloured	0.269***	0.185***	0.0984***	0.354***
	(0.0652)	(0.0685)	(0.0248)	(0.0232)
Asian/Indian	0.256	0.265	0.427***	0.526***
	(0.182)	(0.214)	(0.0630)	(0.0451)
White	0.940***	0.661***	0.832***	0.801***
	(0.0920)	(0.0982)	(0.0366)	(0.0256)
Up to primary	0.0554	0.0595	0.289***	0.242***
	(0.239)	(0.244)	(0.0574)	(0.0493)
Up to secondary	0.536**	0.619***	0.773***	0.620***
	(0.234)	(0.238)	(0.0546)	(0.0467)
Matric	0.918***	0.937***	1.196***	1.023***
	(0.236)	(0.240)	(0.0574)	(0.0471)
Tertiary	1.665***	1.577***	1.933***	1.745***
	(0.235)	(0.239)	(0.0562)	(0.0486)
Constant	6.498***	7.013***	6.846***	7.388***
	(0.254)	(0.261)	(0.0680)	(0.0560)
Observations	3,020	2,067	9,548	19,694
R-squared	0.266	0.255	0.379	0.274

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Summary

While there are statistically significant differences in the distributions of observable characteristics, most of these differences are of little practical significance or can be explained by differences in sampling, survey design or plausibly reflect the impact of the Covid-19 lockdown. Bias, due to non-random attrition across the waves of NIDS, is evident in a higher employment rate than in Stats SA's labour force data. The question around the individual's child support grant receipt was clearly misinterpreted by many respondents and researchers should only use the reports of household level receipt in their analyses. In general, relationships between key variables (e.g. the correlates of earnings, the age gradient in employment) accord with previous research and are similar across datasets. These findings support the use of NIDS-CRAM to explore the unfolding impact of lockdown on individual and household wellbeing.

4. Household income

Missing data

The high rate of missing data for household income in NIDS-CRAM is compared to other waves of NIDS in *Table 7*, including the recently completed second wave of NIDS-CRAM. In all other waves, respondents who did not initially provide an amount were given the option of providing a bracket response. This option can dramatically decrease non-response rates as seen in the comparison between NIDS-CRAM Wave 1 and Wave 2. Non-response rates fell from 38 percent to 16 percent with 21 percent of NIDS-CRAM Wave 2 respondents answering in brackets. In general, non-response rates for income in NIDS-CRAM do not compare favourably with NIDS, although there is substantial variation in missing data across the NIDS waves. Non-response rates ranged from 21 percent in NIDS Wave 1 to three percent in Wave 5.

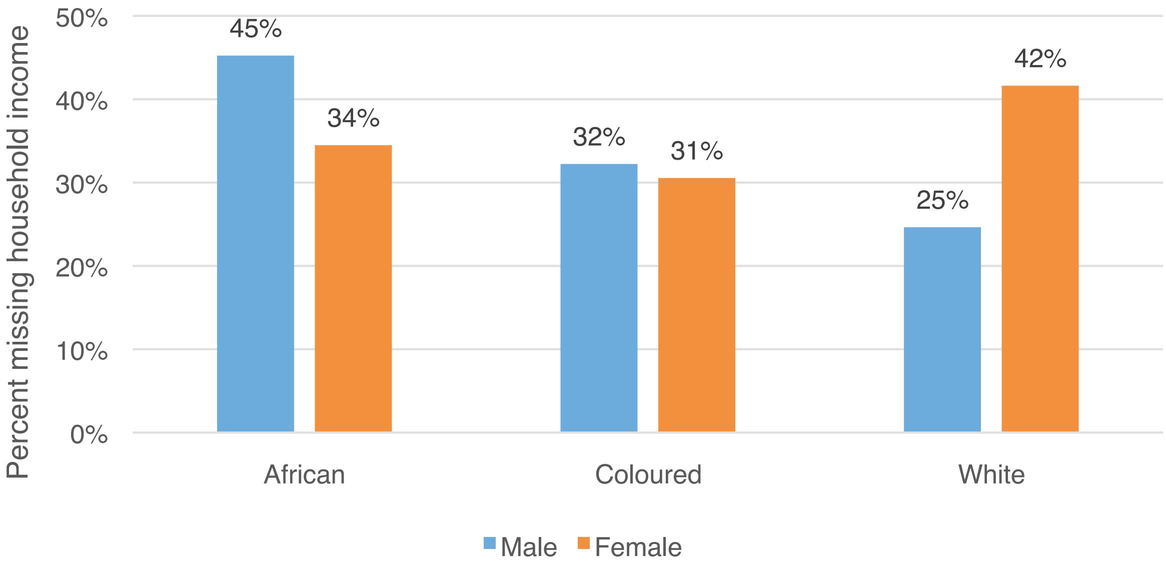
Table 7. Household income missing data

	NIDS-CRAM Wave 1	NIDS-CRAM Wave 2	NIDS Wave 5 – 18plus	NIDS Wave 1
Income in brackets	NA	21.3%	5.0%	7.3%
Income missing	37.7%	15.6%	3.3%	21.4%

Note: Unweighted, at the individual level and restricted to those aged 18 and older for NIDS.

The high rate of missing data will likely distort estimates of the distribution of income and complicate the analysis of poverty and inequality dynamics. The extent of non-response bias depends not only on prevalence but on the non-response mechanism. If non-response is completely random, then estimates of household income will be imprecise but not biased. *Figure 9* summarises non-response rates by race and sex, clearly showing that non-response is not completely random. The highest prevalence of non-response is among African males and White females, with the lowest non-response amongst White males.

Figure 9. Missing household income by race and sex



There are several non-competing hypotheses for the higher rate of non-response in NIDS-CRAM than in previous waves of NIDS including sample and questionnaire design, data collection method and Covid-19 related uncertainty. NIDS-CRAM is a panel of individuals with each individual answering questions about their household. This is in contrast to NIDS, where all members of a household were interviewed and the member(s) most knowledgeable about income and expenditure responded to the household questionnaire. As expected, non-response rates in NIDS-CRAM are lower for those individuals who answered the household questionnaire in NIDS wave 5 (32 percent vs 43 percent). Non-response rates are higher for individuals living in larger households. For each additional household member, the probability of having missing income data increases by around three percentage points. Individuals who moved when lockdown was announced are five percentage points more likely to have missing household income.

Individuals facing greater uncertainty due to the economic impact of lockdown may be less able to provide a household income figure. Following Jain et al. (2020), an individual's working status in April is assigned to one of four categories – working, paid leave, unpaid leave and not working. Those in the unpaid leave category represent those who report having a job, or a job to return to, but working zero hours and receiving no pay. These furloughed workers face a high degree of uncertainty that may explain their higher rate of missing household income data evident in *Figure 10*.

Figure 10. Missing household income by working status

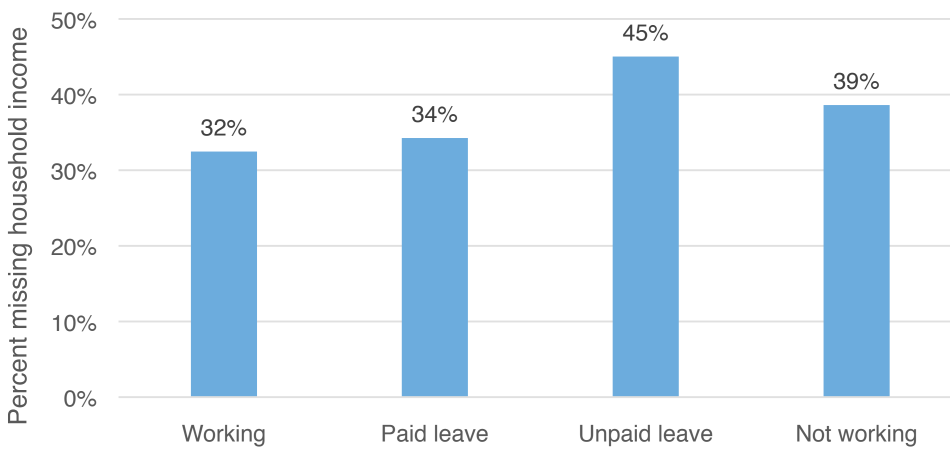
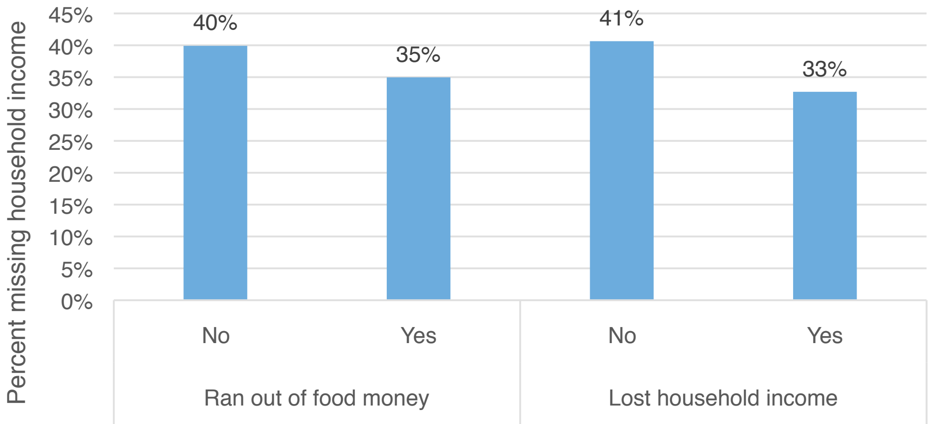


Figure 11 presents non-response rates by two variables that indicate whether the household has experienced economic hardship during lockdown – an indicator that the household ran out of food to buy money and an indicator of household income loss. Interestingly, individuals in households experiencing these events are less likely to have missing household income. One possible explanation is that they are more acutely aware of the sources and level of income in the household.

Figure 11. Missing household income by food insecurity and loss of income



Correlates of having missing household income are summarised in Appendix *Table A2* that presents results from a probit regression of an indicator that household income is missing on a range of variables including those highlighted above. This cursory examination of the correlates of household missing data, makes clear that non-response is non-random and will likely bias estimates. Adjustments to address item non-response bias typically involve imputation and previous waves of NIDS have provided imputed values for key variables such as household income and expenditure using a combination of logical rules and predicted values from regression models. NIDS follows a rule of only imputing values for variables where the non-response rate is less than 40 percent. Household income in NIDS-CRAM is then possibly a candidate for imputation.

However, the suitability of such imputations depends crucially on the plausibility of income data being Missing at Random (MAR) and the quality of the non-missing data underlying the imputation model. MAR means that the propensity of missing values is related to observed data but, conditional on such observed data, is not related to the missing income data. The extensive information gathered prior to NIDS-CRAM together with non-missing Wave 1 variables available for use in imputation models lends justification to the assumption of MAR. The quality of the non-missing household income data in NIDS-CRAM is considered below.

Quality of non-missing data

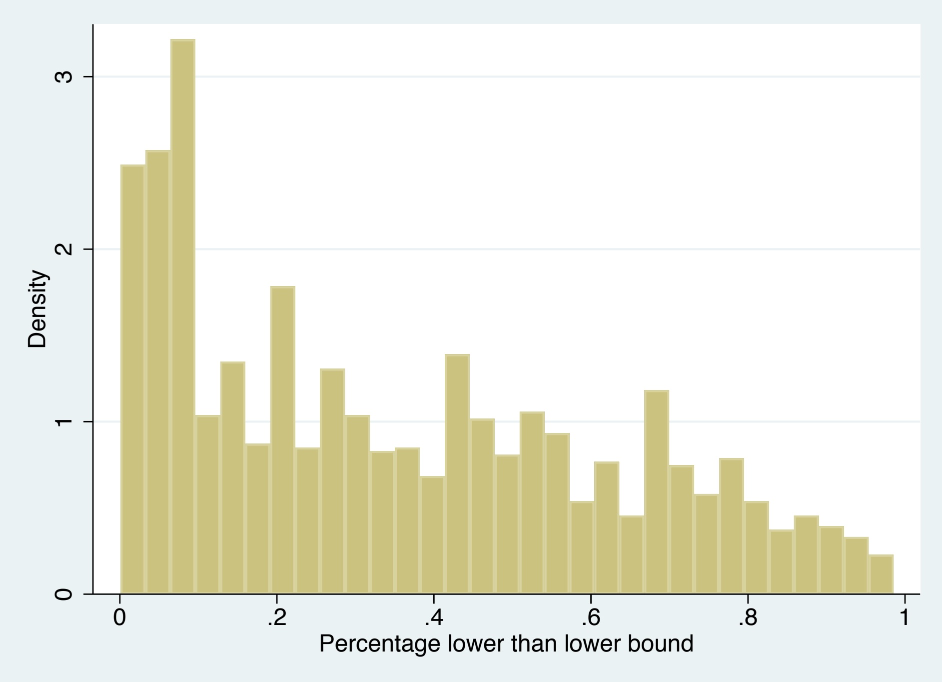
Following Jain et al. (2020), a lower bound for household income is created by combining individual earnings after tax and individual grant income with household grant income based on the reported number of child support grants and state old age pensions received by the household[9]. There are 4,408 individuals with non-missing household income, 3,835 of whom have a non-zero lower bound household income. Of these, the lower bound is higher than the reported household income for 46 percent of individuals[10]. The extent of this discrepancy is summarised in *Figure 12* for those with non-zero household income[11]. Just over a quarter (27 percent) of discrepancies are below 10 percent but around half of the discrepancies are in excess of 30 percent (i.e. reported household income is 30 percent lower than the calculated lower bound income). For the top decile of discrepancies, the lower bound ranges from 4 to 69 times reported household income.

9 If an individual reports receiving the state old age pension, then the number of state old age pensions received by the household is reduced by one before calculating household grant income. Similarly, for the child support grant.

10 Replacing the number of child support grants by the number of children in cases where the number of grants exceed the number of children, and similarly for the number of state old age pensions, reduces the lower bound but it is still greater than reported household income for 43 percent of those individuals with a non-zero lower bound.

11 229 individuals have a non-zero lower bound and report zero household income.

Figure 12. Comparison of reported income to calculated lower bound - individuals with non-zero reported income below lower bound

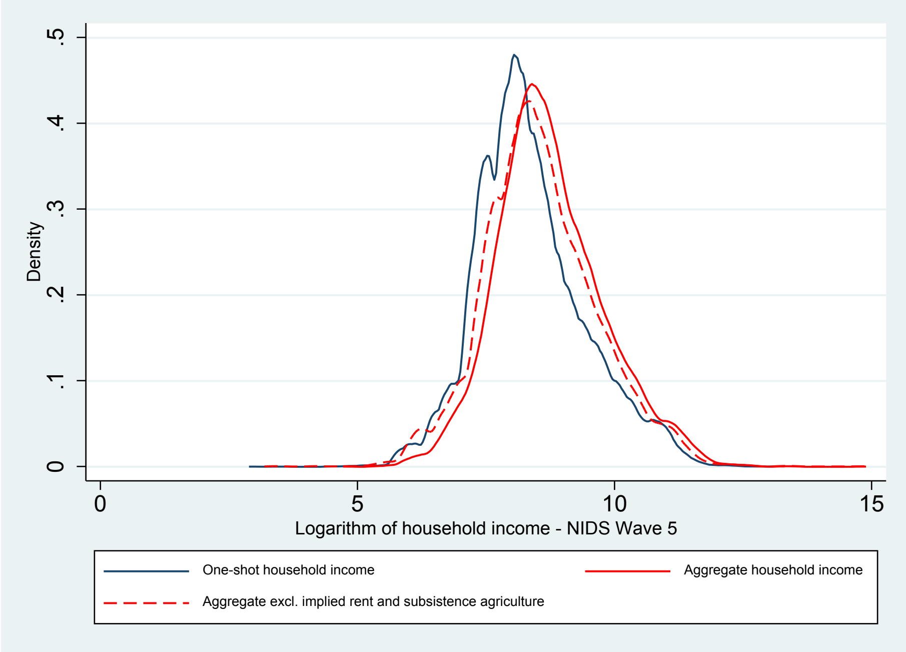


The second column in Appendix *Table A2* presents results from a probit regression of an indicator that the calculated lower bound exceeds reported household income on a range of respondent characteristics. The results show that under-reporting of income is systematically related to gender, race, age, employment status amongst others.

Interestingly, a comparison with NIDS Wave 5 reveals that under-reporting of household income is not only a concern for NIDS-CRAM. A similar analysis to above, found that 42 percent of individuals aged 18 and older in NIDS had a recorded household income that was lower than a constructed lower bound based on just their individual earned and grant income and other grant income in the household.

Jain et al. (2020) point to further indications within NIDS Wave 5 that the one-shot household income question, similar to that asks in NIDS-CRAM, is likely biased downwards. In addition to this one-shot measure, previous waves of NIDS also provide household income aggregated across all the individual income sources in the individual questionnaires. *Figure 9* shows the distribution of the logarithm of household income in Wave 5 for both the one-shot and the aggregate measures. The distributions indicate that incomes reported for the one-shot question tend to be lower than the aggregated income. The one-shot question is preceded by a set of questions asking whether the household received income from various sources, which would hopefully prime the respondent to think of these potential sources in answering the household income question. These sources align with most of the components of the aggregate measure but do not include imputed rental for owner-occupied housing and value of own production consumed. A second aggregate measure with these components removed is shown by the dotted line. Although the distribution shifts to the left, one-shot income still tends to be lower than the adjusted aggregate measure. To the extent that one believes that the aggregate measure is more accurate, the distributions in *Figure 13* suggest that household income reported in NIDS-CRAM will be biased downwards.

Figure 13. One-shot versus aggregate household income – NIDS Wave 5



Summary

Although the rate of missing household income data falls within the 40 percent threshold for NIDS to potentially be used to generate imputed values, the analysis above gives reason for caution. The one-shot question does not appear to be adequately capturing total income in many households in both NIDS and NIDS-CRAM. The danger of providing officially imputed household income data is that many users will accord such data a degree of certainty that is unwarranted given the high prevalence of missing income data and evidence of likely measurement error in non-missing income data. Household income imputations are better left to individual analysts who need to carefully consider the limitations of any imputation method they employ.

5. Conclusion

Measured against previous waves of NIDS and other South African household surveys, the item non-response rates for the NIDS-CRAM telephone survey are comparable to those from face-to-face interviews. Similar to these face-to-face household surveys, non-response rates are highest for income and earnings. Comparisons of key descriptives against appropriate benchmark datasets suggest that the data are broadly representative of the national situation. Estimates of the number of grant recipients show variation between datasets but are mostly within a reasonable range of official SASSA numbers. Moving beyond levels or counts, associations between key outcomes and demographic characteristics accord with expectations and are fairly consistent across datasets. Taken as a whole, the general results are supportive of the NIDS-CRAM data being valid and not subject to major non-response bias. Still, analysts and policy makers should be very cautious with analyses that rely on household incomes during the lockdown.

REFERENCES

Casale, D. & Posel, D. (2020) Gender & the early effects of the COVID-19 crisis in the paid & unpaid economies in South Africa.

Ingle, K., Brophy, T., & Daniels, R. (2020) National Income Dynamics Study – Coronavirus Rapid Mobile Survey (NIDS-CRAM) panel user manual.

Jain, R., Budlender, J., Zizzamia, R., & Bassier, I. (2020) The labour market and poverty impacts of COVID-19 in South Africa.

Kerr, A., Ardington, C., & Burger, R. (2020a) Sample design and weighting in the NIDS-CRAM survey

Kerr, A., Burger, R., Ardington, C., & Spaull, N. (2020b) Representivity in NIDS-CRAM

Posel, D. & Casale, D. (2020) Who moves during times of crisis? Mobility, living arrangements and COVID-19 in South Africa.

Ranchhod, V. & Daniels, R. C (2020) Labour market dynamics in the time of COVID-19.

SASSA (2020) April 2020 Social Grant Payment Report https://www.sassa.gov.za/statistical-reports/Documents/April%202020%20social%20grant%20payment%20report%2011%20June%202020.pdf

Van der Berg, S., Zuze, L., & Bridgman, G. (2020) Coronavirus, Lockdown and Children: Some impacts of the current crisis in child welfare using data from NIDS-CRAM Wave 1.

Wills, G., Patel, L., Van der Berg, S., & Mpeta, B. (2020) Household resource flows and food poverty during South Africa's lockdown: Short-term policy implications for three channels of social protection.

Appendix

Table A1. Item non-response - variables with non-response rates in excess of 5 percent

Description	Variable	% missing	N (including missing)
How much did you make in profit in April (brackets)?	w1_nc_emsinc_apr_brac	43.7%	71
Total household income after tax in April	w1_nc_hhinc	37.7%	7072
How much was your last fortnightly take home pay in April (brackets)?	w1_nc_em2wksinc_apr_brac	35.3%	17
How much was your last monthly take home pay in April (brackets)?	w1_nc_eminc_apr_brac	31.7%	325
Business average work week (days) in April	w1_nc_embusdays_apr	31.0%	393
Does anyone in the household receive a child support grant?	w1_nc_incchld	28.4%	303
How much was your last weekly take home pay in April (brackets)?	w1_nc_emwkinc_apr_brac	28.2%	39
How much was your take-home pay or profit in the month of February (brackets)?	w1_nc_eminc_feb_brac	26.9%	755
How much was your take-home pay or profit in the month of February?	w1_nc_eminc_feb	22.2%	3408
Does anyone in the household receive an old age pension grant?	w1_nc_incgovpen	20.8%	279
How much was your last daily take home pay in April?	w1_nc_emdayinc_apr	19.2%	130
How much was your last monthly take home pay in April?	w1_nc_eminc_apr	18.8%	1727
How much was your take-home pay or profit in April?	w1_nc_emsinc_apr	18.1%	393
How much was your last weekly take home pay in April?	w1_nc_emwkinc_apr	16.9%	231
Business average work day (hours) in April	w1_nc_embushrs_apr	14.8%	271
How likely is your business to close in the next three months?	w1_nc_emscls	14.5%	393
How much was your last fortnightly take home pay in April?	w1_nc_em2wksinc_apr	14.0%	121
Do you think you are likely to get the Coronavirus?	w1_nc_cvrsk	12.9%	7073
What are some of the symptoms of Coronavirus?	w1_nc_cvsypt1	10.0%	7073
Average work week (days) in April	w1_nc_emdays_apr	9.2%	2359
In the last 7 days, did children in household access educational content online?	w1_nc_edccon	8.9%	5775
When was the last time you worked?	w1_nc_noemdc	8.6%	970
Do you have any paid activity/job that you will return to in next 4 weeks	w1_nc_emreturn	8.2%	4329
Reason unavailable or unwilling to work in the next 7 days	w1_nc_noemex	7.5%	971
Average work week (days) in February	w1_nc_emdays_feb	7.0%	3408

In the last 7 days, did children in household watch educational TV?	w1_nc_edctv	6.7%	5775
In the last 7 days, did children in household listen to educational radio?	w1_nc_edcrad	6.7%	5775
Why not visit a healthcare facility?	w1_nc_hlnocon1	6.6%	532
Number of residents who are under 7 years of age (don't forget babies)	w1_nc_nou7res	6.5%	5381
Occupational code for usual work	w1_nc_emwrk_isco_c	6.4%	2359
How often are you paid?	w1_nc_empay_freq	6.4%	2359
In April, did your business pay employee salaries?	w1_nc_emspaysal	5.9%	135
In the last 7 days, did children in household use their school books?	w1_nc_edcbook	5.6%	5775
Respondent's main form of work	w1_nc_emtyp	5.5%	2753

Table A2. Propensity for non-response on household income and for reporting household income that is lower than the calculated lower bound

	Household income missing	Lower bound > household income
Female	-0.266***	-0.229***
	(0.0333)	(0.0462)
Coloured	0.0242	-0.275***
	(0.0734)	(0.0940)
Asian/Indian	-0.202	-0.508**
	(0.159)	(0.213)
White	0.167**	-0.794***
	(0.0832)	(0.138)
Age	-0.00685***	0.00431***
	(0.00117)	(0.00161)
Employed February	-0.152***	-0.0908*
	(0.0392)	(0.0526)
Employed April	-0.0511	0.714***
	(0.0427)	(0.0574)
Household size	0.0631***	0.0361***
	(0.00549)	(0.00780)
Completed matric	0.00738	-0.145***
	(0.0353)	(0.0472)
Receives individual grant	-0.0935**	0.563***
	(0.0426)	(0.0549)

Household receives grants	0.00820	1.089***
	(0.0395)	(0.0560)
Ran out of money for food	-0.177***	0.0652
	(0.0325)	(0.0435)
Eastern Cape	-0.164*	-0.239**
	(0.0875)	(0.110)
Northern Cape	-0.00438	-0.347***
	(0.0886)	(0.112)
Free State	-0.0755	-0.518***
	(0.0975)	(0.123)
KwaZulu-Natal	0.132*	-0.124
	(0.0804)	(0.104)
North West	0.198**	-0.502***
	(0.0960)	(0.128)
Gauteng	0.0687	-0.262**
	(0.0819)	(0.107)
Mpumalanga	0.107	-0.535***
	(0.0902)	(0.119)
Limpopo	0.364***	-0.313***
	(0.0889)	(0.118)
Constant	-0.167*	-1.239***
	(0.101)	(0.138)
Observations	6,782	4,294

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

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