International Journal of Social Research Methodology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tsrm20

Coverage bias: the impact of eligibility constraints on mobile phone-based sampling and data collection

Helen Suich, Mandy Yap & Trang Pham

To cite this article: Helen Suich, Mandy Yap & Trang Pham (2022) Coverage bias: the impact of eligibility constraints on mobile phone-based sampling and data collection, International Journal of Social Research Methodology, 25:6, 849-860, DOI: 10.1080/13645579.2021.1957310

To link to this article: https://doi.org/10.1080/13645579.2021.1957310

Published online: 04 Aug 2021.

INTERNATIONAL JOURNAL OF SOCIAL RESEARCH METHODOLOGY
2022, VOL. 25, NO. 6, 849–860
https://doi.org/10.1080/13645579.2021.1957310

Routledge
Taylor & Francis Group

Coverage bias: the impact of eligibility constraints on mobile phone-based sampling and data collection

Helen Suich [a], Mandy Yap [b] and Trang Pham [a]

[a]Crawford School of Public Policy, Australian National University, Canberra, Australia; [b]Centre for Aboriginal Economic Policy Research, Australian National University, Canberra, Australia

ABSTRACT

This paper uses Individual Deprivation Measure data from Indonesia and South Africa to demonstrate the effects of coverage bias associated with mobile phone-based sampling and data collection approaches that restrict sampling frames to those who own or have access to a mobile phone – a increasingly common method. Analysis of this data demonstrates substantial differences across multiple living standards indicators and demographic characteristics between those who own a mobile phone, individuals with access to a phone and those who neither own nor access mobile phones. This paper highlights the implications of such sampling approaches, and calls for the explicit consideration of the limitations of inferences that can be drawn from qualitative and quantitative analyses because of these substantive differences.

KEYWORDS

Indonesia; South Africa; sampling and data collection; coverage bias; multidimensional poverty; mobile phones

Introduction

The use of mobile phone-based data collection approaches has increased markedly since the outbreak of the COVID-19 pandemic, in order to mitigate the potential risks associated with face-to-face data capture. These approaches typically restrict sampling frames to individuals or households who own or can access a mobile phone and exclude those who cannot. Using data from two districts in South Sulawesi, Indonesia and a national sample in South Africa, collected face-to-face before the pandemic, which used sampling frames based on national censuses and surveys, we show that there are substantial and important differences between those who own, those who access, and those who neither own nor access mobile phones across a range of living standards outcomes and demographic characteristics. Analyses of data collected via mobile phones are likely to miss capturing the full range of experiences because of this coverage bias, regardless of whether the data collected is qualitative or quantitative, and irrespective of the specific instruments used. Researchers should be fully cognizant of the important implications for the generalisability of analyses undertaken and conclusions reached when using mobile phone-based survey data, particularly in countries with relatively low mobile phone penetration.

One of the consequences of the COVID-19 pandemic, beginning in early 2020, has been a rapid change in data collection approaches to ensure the health and safety of field workers and respondents and as a result of the travel and movement restrictions imposed in many countries. Following a short-lived shift to physically distanced data collection in which interviewers maintained their physical distance and met outside when interviewing (IDinsight, 2020), many data collection efforts in low and middle income (LMICs) countries were paused because of the health risks (Bhajibhakare et al., 2020; Hughes & Velyvis, 2020). However, as time passed, researchers have adapted to

CONTACT Helen Suich helen@helensuich.net PO Box 698, Dickson ACT 2602, Australia

circumstances, largely by shifting from traditional in-person to mobile phone-based data collection approaches. These approaches enabled many researchers to resume activities planned prior to the outbreak of the pandemic, and to undertake new research, including that focused on the understanding the health, economic and social impacts of the pandemic (Chikoti et al., 2020; Spaull et al., 2020; Srinivasan & Clifton, 2020).

The focus of this paper is on the implications of mobile phone-based data collection approaches, where the first criteria – though implicit – for individuals' eligibility to participate in a mobile phone-based interview or survey is their ownership of, or access to, a mobile phone. The consequent exclusion of individuals without access to mobile phones has significant implications for the inferences that can be made from qualitative and quantitative research using such approaches.[1]

These implications may, at first glance, seem to be most relevant to the discussion of the analysis of quantitative data, which more frequently aims to be representative and with generalisable results. However, they are of equal relevance to the collection and analysis of qualitative data. Regardless of the type of data collected, the experiences of those who do not own or have access to mobile phones will not be collected and will remain neglected and poorly understood, should only those people with mobile phone access be eligible to participate in data collection.

Digital data collection

In this paper, we use the term digital data collection to refer to the practice of using digital technologies, primarily mobile phones, as the basis for sampling and capturing primary data from respondents, which can be either interviewer-assisted or self-administered.

Traditionally, research using digital data collection involved two initial steps. First, researchers selected a sample from a sampling frame using any appropriate means, such as through national statistics offices or from a listing of households in the area of enumeration prior to data collection. Once the sample was drawn, researchers then arranged initial face-to-face contact with sampled individuals, and where they were found not to own the device needed for data collection, researchers could distribute devices and/or could provide phone or internet credit where necessary (Christensen et al., 2011; Croke et al., 2012; Van Heerden et al., 2014; World Bank, 2017). This approach was undertaken in recognition that, while digital penetration has increased rapidly, there remains a substantial minority of people in many countries without access to mobile phones and other digital technologies. The second step was important to ensure individuals without technology access were also included in the data collection, thus reducing coverage bias.

However, exacerbated by the COVID-19 pandemic and associated lockdowns, in many cases it is often not possible to conduct the second step (i.e. the face-to-face visits to determine technology ownership or access). Thus, digital data collection strategies since the onset of the COVID-19 pandemic have typically required the use of a sampling frame consisting of a list of mobile phone numbers.[2] These sampling frames are constructed from a list of mobile phone numbers of participants from existing surveys (Kerr et al., 2020), by random digit dialing, purchasing/accessing of lists of active phone numbers (IPA, 2020), or through snowball sampling.

An important aspect of the empirical literature on the use of digital data collection compares traditional data collection methods with those using mobile phones (Clark et al., 2008; Dodge & Chapman, 2018; Escobal & Benites, 2013; Gibson et al., 2017; Klevens et al., 2011; Kreuter et al., 2008; Mercader et al., 2017; O'Toole et al., 2009; Tijdens & Steinmetz, 2016). Others examine aspects of the digital interface design (Herzing, 2020) or assess behavior changes arising from information delivered in person or via the web (Neuenschwander et al., 2013). Greenleaf et al. (2017) attempted to synthesize the evidence in LMICs of mode comparison studies, but was unable to draw conclusions because of insufficient information, though broadly there were few differences between different data collection modes.

Of particular relevance to the practicality of mobile phone-based data collection is mobile phone ownership and access, which has increased rapidly in recent years (GMSA, 2020a, 2020b; ITU, n.d.). There is a body of literature that examines the digital divide – those with and without access to the internet – and to a lesser extent the 'mobile phone divide', focusing on determining the proportions of the population who do and do not have access to these technologies. It compares, at a macro level, the demographic characteristics of those with access to mobile phones or internet connections and those without (e.g. Vimalkumar et al., 2020), identifies the geographic locations of those experiencing the digital divide (Leidig & Teeuw, 2015; Song et al., 2020; Sujarwoto & Tampubolon, 2016) or uses mobile phone coverage and related data as a proxy to map poverty (Šćepanović et al., 2015; Schmid et al., 2017; Steele et al., 2017).

Few empirical studies examining the mobile phone divide extend beyond presenting comparisons of demographic characteristics (e.g. gender, age, rural/urban location) of those with and without access, though some do make associations between mobile phone ownership and improved economic outcomes – greater wealth or income (Mubarak et al., 2020) or improved food security (Tankari, 2018).

Utilizing mobile phone ownership as a way to construct the sampling frame and for data collection is likely to over-represent certain groups, because of their increased likelihood of owning a mobile phone (L'Engle et al., 2018; Leo et al., 2015); as is the case for internet-based surveys (Tijdens & Steinmetz, 2016). However, while studies that use mobile phone ownership for sampling often note that their studies exclude those who do not own or have any access to a mobile phone, they rarely describe explicitly or in detail the implications of this coverage bias for the conclusions derived from their analysis.

Those studies that do address coverage bias propose two main approaches to reducing the bias associated with mobile phone-based sampling and data collection. The first is to broaden the eligibility criteria from mobile phone ownership to ownership *or* access – by interviewing individuals other than the owner of the sampled mobile phone number (Himelein et al., 2020). This broadens coverage to those with access to phones (in addition to those who own them), but continues to exclude those who cannot access a mobile phone. The second approach is to use post-stratification weights on the (quantitative) survey data, to be more demographically representative of the general population (L'Engle et al., 2018; Labrique et al., 2017; Larmarange et al., 2016; Peytchev et al., 2011). This latter approach may improve the estimates made by improving the demographic representation of the sample, but implicitly assumes there are no differences between those with and without access to mobile phones, and therefore does not address coverage bias.

What remains lacking in the empirical literature is an explicit focus on the implications of sampling based on the ownership of mobile phones, and the effects of this coverage bias on generalisability and inference. This paper uses recently collected datasets from South Sulawesi, Indonesia and South Africa to demonstrate that there are fundamental differences in demographic characteristics and a wide range of living standards outcomes between three groups: (i) individuals who own a mobile phone; (ii) individuals who can access a mobile; and (iii) those who neither own nor can access a mobile phone. We argue that research conducted using mobile phone-based sampling and data collection should explicitly acknowledge the limitations of their analyses, particularly with respect to generalisability and to inferences to the broader population.

While this paper is focused on the coverage bias that arises from the use of sampling frames derived from mobile phone numbers, non-response bias introduced either at the sampling stage (i.e. individuals not responding to invitations) or throughout data collection (i.e. individuals refusing to answer certain questions) can also result in key experiences of the sampled population missing, which could lead to further measurement bias.

Data sources and description

The data used in this paper were collected in South Sulawesi, Indonesia and South Africa, in 2018 and 2019 respectively, under the auspices of the Individual Deprivation Measure (IDM) program,

an individual-level gender sensitive measure of multidimensional poverty. The two studies aimed to test the revised IDM measure at different scales – regional and national. The IDM was originally developed to formulate a just and justifiable measure of poverty, grounded in the experiences and priorities of poor men and women, and measuring things known to be important to gender equity.

The Indonesian data was collected from two districts in the province of South Sulawesi, Indonesia. A stratified sampling strategy was employed to ensure district-level rural and urban representation. A total of 5,698 individuals were surveyed from 2,186 households across 216 villages (SurveyMeter, 2018). Data collection was undertaken over two months in 2018 by SurveyMeter, an Indonesian survey company, and the demographic characteristics can be seen in Table 1.

After removing two records with missing demographic information and one refusal to respond for the mobile phone ownership and access question, the resulting dataset for analysis were 5,695 individuals. For brevity, this data is referred to as the Indonesian data throughout the paper.

The South African data is from a national-level sample, with a total of 8,652 individuals from 3,811 households. Enumeration areas were randomly selected across all nine provinces, and stratified by rural/urban location. Dwellings were identified using satellite imagery, and then randomly sampled. Attempts were made to interview all eligible individuals in each household in each sampled dwelling, ensuring a minimum provincial sample size of 800 individuals (ikapadata Pty Ltd, 2019). Data collection was undertaken over five months in 2019 by ikapadata, a Cape Town based survey company. The demographic characteristics of the South African sample are also shown in Table 1.

In both studies, eligible individuals were all individuals residing in a sampled dwelling, aged 16 years and older, who were able to communicate for themselves and competent to give informed and ongoing consent. Virtually identical survey tools were used in both studies and all interviews were conducted face-to-face using CAPI. Interviews were conducted in privacy, by an interviewer of the same gender as the respondent. Respondents in Indonesia were interviewed in Bahasa Indonesia, and in South Africa were able to choose to be interviewed in their preferred language – any of the 11 official spoken languages of the country.

Mobile phone ownership was self-reported by each individual in response to the question 'did you own or are you allowed to use a cell/mobile phone?'. The 'own' category used in the analyses below refers to individuals who reported owning a mobile phone, either by themselves or jointly with others. The 'access' category refers to individuals who stated they had access to, or were allowed to use, a mobile phone, though no information was collected about who owned the phone they could access. Individuals classified as having 'none' stated that they did not own and had no access to a mobile phone. Demographic characteristics by mobile phone ownership are shown in Table 1.

In Indonesia, in 2017, 50.3% of individuals in the two districts of South Sulawesi sampled in the IDM study reported they owned a mobile in the previous three months, and as many as 70.6% of individuals reported they used a mobile phone in the previous three months (i.e. they had access to a mobile phone) (Central Bureau of Statistics, 2018). While not directly comparable due to different questions about ownership and access, the IDM data for mobile ownership does align with figures available from Indonesia's official statistics. 89.5% of South African households reported having a functional mobile connection in their dwellings in 2018, and a further 7.1% with a functional landline and mobile (Statistics South Africa, 2019). These data do broadly align with that of the IDM South Africa study, but are measured at the household rather than individual level, so they are difficult to compare to those collected in the IDM South Africa study.

Indicator selection

A number of indicators from the two datasets are presented below, selected on the basis that they illustrate a range of different living standard outcomes, including non-financial indicators which are

Table 1. Demographic characteristics by mobile phone ownership and overall, Indonesia and South Africa.

	Overall (n)	Overall %	Male	Female	16–24	25–64	65+	Rural	Urban	Without disabilities	With disabilities (strict)	Primary schooling or higher	No schooling
Indonesia (South Sulawesi)													
Own	3382	59.4	63.5	56.0	75.3	59.3	20.9	55.9	73.7	61.5	36.1	64.8	27.9
Access	814	14.3	11.4	16.6	9.97	15.1	17.8	15.5	9.4	13.8	19.8	12.8	23.2
None	1499	26.3	25.1	27.3	14.7	25.6	61.2	28.6	16.9	24.7	44.1	22.4	48.9
Total cases	5695	-	2546	3149	1113	4128	454	4586	1109	5221	474	4863	832
%	-	100	44.7	55.3	19.5	72.5	8.0	80.5	19.5	91.7	8.3	85.4	14.8
South Africa													
Own	7260	83.9	81.4	85.8	78.7	87.4	74.8	87.2	82.2	84.8	76.2	85.1	67.0
Access	774	8.9	9.6	8.5	12.7	7.0	11.9	6.6	10.2	8.7	11.3	8.7	13.7
None	618	7.1	9.0	5.7	8.6	5.5	13.3	6.3	7.6	6.5	12.5	6.2	19.2
Total cases	8652	-	3738	4914	2101	5586	965	3005	5647	7764	888	8092	531
%	-	100	43.2	56.8	23.4	64.6	11.2	34.7	65.3	89.7	10.3	93.8	6.2

not necessarily associated with mobile phone ownership. In addition, these indicators are likely to have been (often negatively) impacted by the COVID-19 pandemic and associated lockdowns.

The physical health indicator assesses whether people reported a long-term health condition and/or an illness related to unclean cooking fuel smoke, while the psycho-social health indicator assesses the frequency of experiences of anxiety and/or depression.

The food security indicator used determines an individuals' experience of compromising the quality and/or quantity of food eaten, due to a lack of financial or other resources. There are likely to be interactions between food security and physical and psycho-social health, though available data does not yet reflect these interactions.

Access to handwashing facilities is included because of the general hygiene and health implications for individuals without such facilities in following recommended COVID-19 hygiene practices. Handwashing facilities are defined as a having a place to wash hands at home with sufficient water and soap or soap substitutes. Those without handwashing facilities lack one or all of these elements.

The indicators of functional literacy and functional numeracy are included because they capture the challenges for individuals accessing and understanding information (including COVID-related information) and because of the implications of illiteracy for individuals' ability to self-administer mobile data collection. Respondents are considered to be functionally literate if they completed simple tasks demonstrating they could read and write adequately in an official language of their country. Individuals are considered to be functionally numerate – at a very basic level – if they were able to correctly complete two sums covering a simple addition/subtraction calculation and a multiplication and division calculation.

The labor force participation indicator categorizes respondents of all age groups as either in the labor force (i.e. working or unemployed), or not. This operational definition means these labor force statistics are different from official statistics where only individuals of working age are included.[3] Those classified as employed were engaged in any activity to produce goods or services for pay or profit, whether in the formal or informal sector, or for own production (e.g. subsistence farming activities), during the seven days prior to the survey. Those classified as unemployed were not in employment but available to take up work and/or conducting employment-seeking activities during this reference period. Those classified as not in the labor force by choice include those who reported that they did not want or need to work, individuals in full time education and those who had retired. Those categorized as not in the labor force due to exclusion are those who reported that they had given up looking for work because it was too hard to find, who were too busy with domestic/caring duties, those with an illness/injury or condition preventing them from working or were unable to work for any other reason, and individuals who were not allowed to work.

The indicators on water and fuel collection are included because many poor households in LMICs are dependent on sources of water and fuel outside the home, and because access to water and energy have knock-on effects on hygiene and food security. These collection activities are typically necessary daily and may involve multiple household members, often women and girls.

The indicator on support determines whether individuals require assistance from people outside the home to help meet basic needs such as food, water and shelter. Those categorized as having 'enough support' either do not require any support or they require support but they receive enough to meet their needs, while those categorized as 'not enough support' are individuals whom reported that they needed support, but did not have enough.

Differences in demographics and key living standards outcomes

Overall, much higher rates of mobile ownership and access are reported in South Africa compared to Indonesia (Table 1). In Indonesia, males were more likely to report owning or having access compared to females; a pattern reversed in South Africa. In Indonesia, mobile ownership was

highest amongst the youth (16–24), whilst ownership in South Africa was highest amongst those aged 25–64. Those without disabilities were more likely than those with disabilities to report owning a mobile phone, though the differences between the two groups was smaller in South Africa than in Indonesia. In both countries, mobile ownership and access was higher amongst those who had primary schooling or higher compared to those with no schooling.

The results for each of the key living standards indicators are presented in Table 2. Reported sample sizes in the table vary – they exclude individuals who refused to answer the question(s) making up each indicator – and are unweighted. Significance tests (chi-squared), examining whether living standard outcomes differ for the three mobile phone status were also undertaken, and p-values are reported in the table.

As described in detail below, Table 2 demonstrates that those who own mobile phones typically reported the best outcomes, followed by those with access, and the worst outcomes achieved by those without a phone. Results discussed below refer to both countries unless otherwise specified.

There were significant differences in physical health outcomes by mobile phone ownership and access – physical health outcomes were worse for those without mobile phones, reporting both health conditions compared to those with mobile phones.

Those without mobile phones were more likely than the other two groups to report feeling anxious and/or depressed at least weekly. Overall, individuals in Indonesia reported somewhat better psycho-social health outcomes than in South Africa. However, for those who do report feeling anxious or depressed, it is more likely to be happening weekly or daily in Indonesia compared to monthly or a few times a year in South Africa. Notwithstanding the lack of appropriate mental health services that might be available, having a mobile phone may be a source of support by enabling connections with friends and family.

Those without mobile phones in both countries were more likely to experience some food insecurity than those with phone ownership or access. The rate of severe food insecurity in Indonesia was around half that in South Africa, where almost half of those without a mobile phone reported severe food insecurity.

Those without mobile phones are less likely to have handwashing facilities than those with mobile phones, and more than one third of each group lack basic handwashing facilities at home, which is likely to have significant negative health implications for these individuals, even in the absence of a global health pandemic.

Approximately 70% of the overall sample in both countries were classified as having basic functional literacy and numeracy. Those without mobile phones were least likely to be classified as functionally literate, and a relatively high proportion of those with access to mobile phones were also not functionally literate. A similar, though more extreme, pattern was observed for functional numeracy, and the share of those who were classified as functionally numerate was much lower in South Africa compared to Indonesia.

Those without mobile phones are less likely to be employed than those with mobile phones, and more likely to be unemployed or excluded from the labor force, though the difference is not large in South Africa.

For both water and fuel collection, those who own mobile phones are least likely to report being responsible for collection in Indonesia, while in South Africa, those with access to mobiles phones are the least likely and those who own them are most likely to be responsible.

As for most other indicators, those without mobile phones reported the worst outcomes with respect to support – they were more likely to need support but not receive enough (even before the pandemic) than other groups.

Discussion and conclusions

While the share of mobile phone ownership has been increasing in recent years, the ownership pattern is quite different between South Sulawesi, Indonesia and South Africa – approximately

Table 2. Living standards indicators by mobile phone ownership, Indonesia and South Africa.

	Indonesia (South Sulawesi)				South Africa			
	Own	Access	None	Overall	Own	Access	None	Overall
Physical health								
No condition	81.0	76.3	74.5	78.6	66.8	62.7	57.9	65.8
One condition only	17.5	22.2	22.0	19.4	28.2	32.4	34.2	29.0
Both conditions	1.4	1.5	3.5	2.0	4.9	4.9	8.0	5.2
n	3380	814	1499	5693	7191	772	603	8577
p-value	0.000				<0.00001			
Psycho-social health								
No anxiety and depression	34.2	43.1	41.0	37.3	19.3	23.3	20.6	19.8
Anxiety and/or depression monthly or a few times a year	44.2	34.2	35.0	40.3	69.8	66.2	63.7	69.1
Anxiety and depression experienced at least weekly	21.6	22.7	24.0	22.4	10.9	10.5	15.7	11.1
n	3381	814	1499	5694	7244	772	606	8622
p-value	0.000				0.00021			
Food security								
Food secure	52.1	44.9	37.9	47.3	38.0	30.2	23.1	36.2
Mild food insecurity	20.7	23.6	24.0	22.0	15.5	10.3	9.2	14.6
Moderate food insecurity	13.2	15.4	16.9	14.5	18.1	19.8	20.7	18.5
Severe food insecurity	14.0	16.1	21.2	16.2	28.4	39.7	46.9	30.7
n	3382	814	1499	5688	7260	774	618	8652
p-value	0.000				<0.00001			
Handwashing facilities								
Those with handwashing facilities	65.4	54.3	51.4	60.1	57.9	56.3	46.1	56.9
Those without handwashing facilities	34.6	45.7	48.6	39.9	42.1	43.7	53.9	43.1
n	3382	814	1499	5695	7260	774	618	8652
p-value	0.000				<0.00001			
Functional literacy								
Functionally literate	86.7	60.2	52.8	74.3	73.6	63.1	47.8	70.8
Not functionally literate	13.3	39.8	47.2	25.7	26.4	36.9	52.2	29.2
n	3319	776	1390	5485	7001	738	573	8322
p-value	0.000				<0.00001			
Functional numeracy								
Functionally numerate	85.5	73.3	62.0	77.6	65.2	50.9	38.8	62.2
Not functionally numerate	14.5	26.7	38.0	22.4	34.8	49.1	61.2	37.8
n	3371	813	1495	5679	6802	688	539	8039
p-value	0.000				<0.00001			
Labour force participation								
Employed	74.0	68.5	61.9	70.1	38.6	22.6	21.5	35.9
Unemployed	9.4	10.6	13.1	10.5	24.4	25.4	24.8	24.6
Not in the labour force – by choice	6.0	4.7	4.5	5.4	25.5	39.0	40.4	27.8
Not in the labour force – exclusion	10.6	16.2	20.5	14.0	11.5	13.0	13.2	11.7
n	3382	814	1499	5695	7224	771	604	8610
p-value					<0.00001			
Water collection								
Responsible	23.2	26.0	32.3	26.0	42.2	32.7	38.1	41.0
Not responsible	76.8	74.0	67.7	74.0	57.8	67.3	61.9	59.0
n	3382	814	1499	5695	7252	774	607	8644
p-value	0.000				<0.00001			
Fuel collection								
Responsible	19.7	34.0	32.7	25.1	32.3	24.9	30.2	31.5
Not responsible	80.3	66.0	67.3	74.9	67.7	75.1	69.8	68.5
	3382	814	1499	5695	7247	772	606	8636
p-value	0.000				0.00011			
Support								
Enough support	92.0	92.1	87.4	90.8	87.6	81.5	77.8	86.3
Not enough support	8.0	7.9	12.6	9.2	12.4	18.5	22.2	13.7

(*Continued*)

Table 2. (Continued).

	Indonesia (South Sulawesi)				South Africa			
	Own	Access	None	Overall	Own	Access	None	Overall
n	3382	814	1499	5695	6969	728	580	8287
p-value	0.000				<0.00001			

26.3% reported neither owning nor having access to a mobile phone, compared to just 7.1% in South Africa.

Demographic differences between those who own, can access or have no access to a mobile phone are apparent. In Indonesia, groups not owning or having access a mobile phone are more likely to be women, those aged 65 years or older, individuals living in rural areas and people with disabilities. In South Africa, men, those over 65 and people with disabilities are more likely to report not owning or accessing a mobile phone.

Importantly, the data presented demonstrates significant and substantial differences in living standards outcomes between the three groups of people who owned, who had access and who did not own or have access to a mobile phone. The pattern of differential outcomes between the three groups is similar in both studies – broadly those who did not own or access a mobile phone had the worst outcomes across the indicators, with better outcomes for those who could access a mobile phone, and the best outcomes for those who owned a mobile phone.

The COVID-19 pandemic has increased the use of data collection approaches that can be implemented remotely, often those that rely on mobile phone numbers for sampling, whether for qualitative or quantitative data collection. However, the coverage bias introduced – particularly in countries with lower mobile phone penetration rates, and when there is no possibility to distribute mobile phones to potential participants who do not already own or access them – has important implications and should be explicitly considered by researchers, regardless of the instruments used, the nature of the data collected and the methodological approaches employed.[4]

Even where data collected using mobile phone-based sampling can be weighted to reflect demographics at the relevant scale, the analysis presented in this paper suggests that the demographic and other differences between the three groups cannot be adjusted for, because there is typically little or no information about these differences between the three groups at each research site. Thus, findings drawn from a sample restricted to individuals who owned a mobile phone (and perhaps those who accessed one) are unlikely to be generalizable, and caution should be used in drawing inferences from analyses of such samples.

The results presented above suggest that digital data collection sampling strategies should be explicitly designed to include those who own mobile phones *and* those who can access them, in order to reduce coverage bias. However, this does not fully resolve the problem, as individuals without access to mobile phones would continue to be excluded without a specific intervention to deliver the relevant digital devices that could facilitate their participation.

Given the use of sampling frames consisting of mobile phone numbers is a practical response to constraints imposed by the pandemic – and is likely to remain an important approach in the future due to its cost-effectiveness – the extent of the coverage bias introduced from such approaches requires further interrogation. This paper calls for research which is specifically designed to improve the understanding of the implications of this coverage bias – and indeed of non-response and other biases associated with this data collection mode which are not covered in this paper.

It is critically important for researchers using mobile phone-based data collection strategies to be mindful that the experiences of those who do not own and cannot access a mobile phone will be excluded from their research. If the excluded views and experiences of those without mobile phones

are substantively different from the general population – as the results presented here indicate – policy responses designed from analyses of such data are likely to be poorly targeted or potentially misguided.

Notes

1. The question of whether digital data collection is appropriate for all types of research is beyond the scope of this paper.
2. Similarly, if these approaches were to be used for landline/fixed telephone data collection, the sampling frame would consist of landline phone numbers, as they have historically been in developed countries.
3. The three age groups were selected to distinguish the youth (16–24), and older individuals, largely those of retirement age (i.e. 65+) from the working age population (25–64).
4. These implications are relevant to both interviewer-assisted and self-administered approaches, though it may be easier to use the former to ensure and verify the inclusion of those who can access but do not own a mobile phone.

Acknowledgments

This research was conducted using data from the Individual Deprivation Measure Program (2016–2020), which was a partnership between the Australian National University and the International Women's Development Agency, with funding from the Australian Department of Foreign Affairs and Trade. The authors acknowledge the contribution of Sharon Bessell and Janet Hunt and the IWDA-IDM team to the IDM survey development and data collection. All views are those of the authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Helen Suich is an independent researcher. Her research focuses broadly on issues of poverty alleviation and rural development, and the evaluation of projects and programs designed to achieve these goals. Helen worked for the Individual Deprivation Measure programme at the ANU and led the South Africa Country Study.

Mandy Yap is a Fellow at the Centre for Aboriginal Economic Policy Research (CAEPR) at the Australian National University. She was recently awarded an Australian Research Council Discovery Early Career Researcher Award (2021–2024). Before returning to CAEPR in 2019, she was employed at the Crawford School of Public Policy working on the Individual Deprivation Measure (IDM), a gender sensitive measures individual deprivation. She co-led the IDM study in two districts in South Sulawesi, Indonesia. Prior to joining the ANU, Mandy worked the National Centre for Social and Economic Modelling. Mandy has an interest in of composite measures of quality of life with a particular focus on methodologies surrounding selection and weighting of composite measures of wellbeing which gives priority to the lived realities and perspectives of communities and individuals on the ground. Since 2013, Mandy has been working in partnership with the Yawuru community in Broome to develop culturally-relevant indicators of Indigenous wellbeing. Mandy is an Australian Endeavour Fellow and obtained her PhD from the ANU in 2017.

Trang Pham works in the area of social policy research, international development, and evaluation. She has consulted for the World Bank and the United Nations Women on the topics of community driven development, poverty reduction and women peace and security. Trang worked at the Individual Deprivation Measure (IDM) program at the Australian National University as a research associate. As a research associate developing a gender-sensitive measure of multidimensional poverty developing a gender-sensitive measure of multidimensional poverty, She is currently a senior evaluator for the New South Wales government.

ORCID

Helen Suich http://orcid.org/0000-0002-5014-8981
Mandy Yap http://orcid.org/0000-0002-0879-0313
Trang Pham http://orcid.org/0000-0002-7841-1518

References

Bhajibhakare, S., Gupta, P., & Kopper, S. (2020). *Resources for adapting to phone surveys during COVID-19*. Poverty Action Lab Blog. https://www.povertyactionlab.org/blog/7-2-20/resources-adapting-phone-surveys-during-covid-19

Central Bureau of Statistics. (2018). *Survei Sosial Ekonomi Nasional (SUSENAS) 2017 (Indonesia national socio-economic survey)*. Central Bureau of Statistics, Indonesia.

Chikoti, L., Vundru, W. D., Fuje, H. N., Ilukor, J., Kanyanda, S. S. E., Kanyuka, M., Kilic, T., Mleme, T., Moylan, H. G., Mvula, B., Mwale, I., Mwalwanda, T., & Yoshida, N. (2020). *Monitoring COVID-19 impacts on households in Malawi: Findings from the first round of the high-frequency phone survey*. World Bank.

Christensen, P., Mikkelsen, M. R., Nielsen, T. A. S., & Harder, H. (2011). Children, mobility and space: Using GPS and mobile phone technologies in ethnographic research. *Journal of Mixed Methods Research*, *5*(3), 227–246. https://doi.org/10.1177/1558689811406121

Clark, M. A., Rogers, M. L., Armstrong, G. F., Rakowski, W., & Kviz, F. J. (2008). Differential response effects of data collection mode in a cancer screening study of unmarried women ages 40-75 years: A randomized trial. *Bmc Medical Research Methodology*, *8*, 10, Article 10. https://doi.org/10.1186/1471-2288-8-10

Croke, K., Dabalen, A., Demombynes, G., Giugale, M., & Hoogeveen, J. (2012). *Collecting high frequency panel data in Africa using mobile phone interviews Policy Research Working Paper no. 6097*. World Bank.

Dodge, N., & Chapman, R. (2018). Investigating recruitment and completion mode biases in online and door to door electronic surveys. *International Journal of Social Research Methodology*, *21*(2), 149–163. https://doi.org/10.1080/13645579.2017.1336596

Escobal, J., & Benites, S. (2013). PDAs in socio-economic surveys: Instrument bias, surveyor bias or both? *International Journal of Social Research Methodology*, *16*(1), 47–63. https://doi.org/10.1080/13645579.2011.648420

Gibson, D. G., Pereira, A., Farrenkopf, B. A., Labrique, A. B., Pariyo, G. W., & Hyder, A. A. (2017). Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: A literature review. *Journal of Medical Internet Research*, *19*(5), e139, Article e139. https://doi.org/10.2196/jmir.7428

GMSA. (2020a). *The mobile economy, Asia Pacific 2020*. GMSA. https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/06/GSMA_MobileEconomy_2020_AsiaPacific.pdf

GMSA. (2020b). *The mobile economy, sub-Saharan Africa 2020*. GMSA. https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/09/GSMA_MobileEconomy2020_SSA_Eng.pdf

Greenleaf, A. R., Gibson, D. G., Khattar, C., Labrique, A. B., & Pariyo, G. W. (2017). Building the evidence base for remote data collection in low- and middle-income countries: Comparing reliability and accuracy across survey modalities. *Journal of Medical Internet Research*, *19*(5), e140, Article e140. https://doi.org/10.2196/jmir.7331

Herzing, J. M. E. (2020). Investigation of alternative interface designs for long-list questions–the case of a computer-assisted survey in Germany. *International Journal of Social Research Methodology*, *23*(6), 639–650. https://doi.org/10.1080/13645579.2020.1723201

Himelein, K., Eckman, S., Lau, C., & McKenzie, D. (2020). *Mobile phone surveys for understanding COVID-19 impacts: Part I sampling and mode*. *Development Impact*. Washington, DC: World Bank. https://blogs.worldbank.org/impactevaluations/mobile-phone-surveys-understanding-covid-19-impacts-part-i-sampling-and-mode

Hughes, S., & Velyvis, K. (2020). *Tips to quickly switch from face-to-face to home-based telephone interviewing*. Mathematica. Princeton, NJ: Mathematica https://www.mathematica.org/commentary/tips-to-quickly-switch-from-face-to-face-to-home-based-telephone-interviewing

IDinsight. (2020). *Data collection practices and recommendations for COVID-19*. IDinsight. https://www.idinsight.org/data-collection-practices-and-recommendations-for-covid-19

ikapadata Pty Ltd. (2019). ANU IDM Survey 2019. Fieldwork report. [Unpublished report prepared for the ANU IDM team].

IPA. (2020). *Remote surveying in a pandemic: Handbook* https://www.poverty-action.org/sites/default/files/publications/IPA-Phone-Surveying-in-a-Pandemic-Handbook.pdf

ITU. (n.d.). *ITU ICT-Eye data portal*. Retrieved September 23, 2020, from https://www.itu.int/net4/ITU-D/icteye/#/

Kerr, A., Ardington, C., & Burger, R. (2020). *Sample design and weighting in the NIDS-CRAM survey Wave 1. National Income Dynamics Study (NIDS) – Coronavirus Rapid Mobile Survey (CRAM)*. NIDS-CRAM.

Klevens, J., Trick, W. E., Kee, R., Angulo, F., Garcia, D., & Sadowski, L. S. (2011). Concordance in the measurement of quality of life and health indicators between two methods of computer-assisted interviews: Self-administered and by telephone. *Quality of Life Research*, *20*(8), 1179–1186. https://doi.org/10.1007/s11136-011-9862-2

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability in CATI, IVR and web surveys. The effects of mode and question sensitivity. *Public Opinion Quarterly*, *72*(5), 847–865. https://doi.org/10.1093/poq/nfn063

L'Engle, K., Sefa, E., Adimazoya, E. A., Yartey, E., Lenzi, R., Tarpo, C., Heward-Mills, N. L., Lew, K., & Ampeh, Y. (2018). Survey research with a random digit dial national mobile phone sample in Ghana: Methods and sample quality. *PLoS ONE*, *13*(1), e0190902. https://doi.org/10.1371/journal.pone.0190902

Labrique, A., Blynn, E., Ahmed, S., Gibson, D., Pariyo, G., & Hyder, A. A. (2017). Health surveys using mobile phones in developing countries: Automated active strata monitoring and other statistical considerations for improving

precision and reducing biases. *Journal of Medical Internet Research*, *19*(5), e121, Article e121. https://doi.org/10.2196/jmir.7329

Larmarange, J., Kassoum, O., Kakou, É., Fradier, Y., Sika, L., & Danel, C. (2016). Feasibility and representativeness of a random sample mobile phone survey in Côte d'Ivoire. *Population*, *17*(1), 119–129. https://doi.org/10.3917/pope.1601.0119

Leidig, M., & Teeuw, R. M. (2015). Quantifying and mapping global data poverty. *PLoS ONE*, *10*(11), e0142076, Article e0142076. https://doi.org/10.1371/journal.pone.0142076

Leo, B., Morello, R., Mellon, J., Peixoto, T., & Davenport, S. (2015). *Do mobile phone surveys work in poor countries? Working paper 398*. CGD.

Mercader, H. F. G., Kabakyenga, J., Katuruba, D. T., Hobbs, A. J., & Brenner, J. L. (2017). Female respondent acceptance of computer-assisted personal interviewing (CAPI) for maternal, newborn and child health coverage surveys in rural Uganda. *International Journal of Medical Informatics*, *98*, 41–46. https://doi.org/10.1016/j.ijmedinf.2016.11.009

Mubarak, F., Suomi, R., & Kantola, S.-P. (2020). Confirming the links between socio-economic variables and digitalization worldwide: The unsettled debate on digital divide. *Journal of Information Communication & Ethics in Society*, *18*(3), 415–430. https://doi.org/10.1108/jices-02-2019-0021

Neuenschwander, L. M., Abbott, A., & Mobley, A. R. (2013). Comparison of a web-based vs in-person nutrition education program for low-income adults. *Journal of the Academy of Nutrition and Dietetics*, *113*(1), 120–126. https://doi.org/10.1016/j.jand.2012.07.034

O'Toole, J. E., Sinclair, M. I., & Leder, K. (2009). Collecting household water usage data: Telephone questionnaire or diary? *Bmc Medical Research Methodology*, *9*, 72, Article 72. https://doi.org/10.1186/1471-2288-9-72

Peytchev, A., Carley-Baxter, L. R., & Black, M. C. (2011). Multiple sources of nonobservation error in telephone surveys: Coverage and nonresponse. *Sociological Methods & Research*, *40*(1), 138–168. https://doi.org/10.1177/0049124110392547

Šćepanović, S., Mishkovski, I., Hui, P., Nurminen, J. K., & Ylä-Jääski, A. (2015). Mobile phone call data as a regional socio-economic proxy indicator. *PLoS ONE*, *10*(4), e0124160. https://doi.org/10.1371/journal.pone.0124160

Schmid, T., Bruckschen, F., Salvati, N., & Zbiranski, T. (2017). Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: Estimating literacy rates in Senegal. *Journal of the Royal Statistical Society Series a-Statistics in Society*, *180*(4), 1163–1190. https://doi.org/10.1111/rssa.12305

Song, Z., Wang, C., & Bergmann, L. (2020). China's prefectural digital divide: Spatial analysis and multivariate determinants of ICT diffusion. *International Journal of Information Management*, *52*. https://doi.org/10.1016/j.ijinfomgt.2020.102072

Spaull, N., Ardington, C., Bassier, I., Bhorat, H., Bridgman, G., Brophy, T., Budlender, J., Burger, R., Burger, R., Carel, D., Casale, D., Christian, C., Daniels, R., Ingle, K., Jain, R., Kerr, A., Köhler, T., Makaluza, N., Maughan-Brown, B., Mpeta, B., … Zuze, L. (2020). *Overview and findings. NIDS-CRAM synthesis report wave 1*. NIDS-CRAM, UCT.

Srinivasan, R., & Clifton, J. (2020). *Gallup keeps listening to the world amid the pandemic*. Gallup blog. Washington, DC: Gallup Inc. https://news.gallup.com/opinion/gallup/316016/gallup-keeps-listening-world-amid-pandemic.aspx

Statistics South Africa. (2019) . *General household survey 2018*. Stats SA.

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., De Montjoye, Y.-A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A. J., & Bengtsson, L. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface*, *14*(127), 20160690. https://doi.org/10.1098/rsif.2016.0690

Sujarwoto, S., & Tampubolon, G. (2016). Spatial inequality and the Internet divide in Indonesia 2010-2012. *Telecommunications Policy*, *40*(7), 602–616. https://doi.org/10.1016/j.telpol.2015.08.008

SurveyMeter. (2018). Implementation report: IDM data collection and data cleaning activities [Unpublished report prepared for the ANU IDM team].

Tankari, M. R. (2018). Mobile phone and households' poverty: Evidence from Niger. *Journal of Economic Development*, *43*(2), 67–84. https://doi.org/10.35866/caujed.2018.43.2.004

Tijdens, K., & Steinmetz, S. (2016). Is the web a promising tool for data collection in developing countries? An analysis of the sample bias of 10 web and face-to-face surveys from Africa, Asia, and South America. *International Journal of Social Research Methodology*, *19*(4), 461–479. https://doi.org/10.1080/13645579.2015.1035875

van Heerden, A. C., Norris, S. A., Tollman, S. M., Stein, A. D., & Richter, L. M. (2014). Field lessons from the delivery of questionnaires to young adults using mobile phones. *Social Science Computer Review*, *32*(1), 105–112. https://doi.org/10.1177/0894439313504537

Vimalkumar, M., Singh, J. B., & Sharma, S. K. (2020). Exploring the multi-level digital divide in mobile phone adoption: A comparison of developing nations. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-020-10032-5

World Bank. (2017) . *Listening to Tajikistan. Household survey: Background, implementation, and methods.*