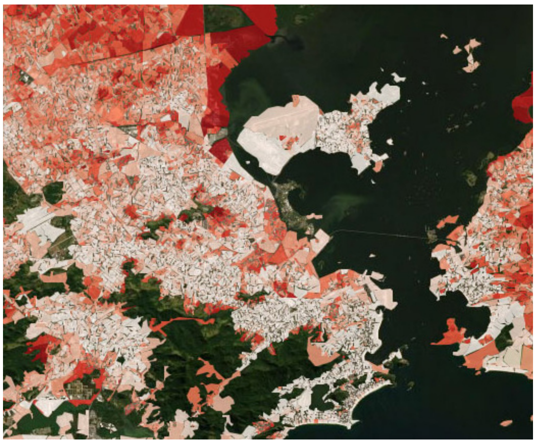
MRC/CSO Social and Public Health Sciences Unit

Developing a small-area deprivation measure for Brazil: Technical report



Mirjam Allik, Dandara Ramos, Marilyn Agranonik, Elzo Pereira Pinto Júnior, Maria Yury Ichihara, Mauricio L Barreto, Alastair H Leyland, Ruth Dundas

2020

FUNDED BY NIHR National Institute for Health Research

UKaid from the British people

Cover Images | Tiles © Esri – Source: Esri i-cubed USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP and the GIS User Community

The cover images show deprivation by census sectors in the city of Manaus (population 2.2 million) in the state of Amazon, North of Brazil (left image) and in Rio de Janeiro (population 6.7 million), Southeast of Brazil (right). The darker red areas show high levels of deprivation and the lighter areas lower deprivation. In Brazil, deprivation is higher in the North, Northeast and in rural areas, such as those surrounding Manaus. The South, Southeast and urban areas, for example many of the seafront neighbourhoods of Rio de Janeiro, are among the least deprived. However, large urban areas are also home to *favelas* and pockets of high deprivation, such as shown by the dark red areas within Rio de Janeiro.

Contents

1 Summary

This report describes the development of the BrazDep small-area deprivation measure for the whole of Brazil. The measure uses the 2010 Brazilian Population Census data and is calculated for the smallest possible geographical area level, the census sectors. It combines three variables – (1) percent of households with *per capita* income $\leq 1/2$ minimum wage; (2) percent of people not literate, aged 7+; and (3) average of percent of people with inadequate access to sewage, water, garbage collection and no toilet and bath/shower – into a single measure. Similar measures have previously been developed at the census sector level for some states or municipalities, but the deprivation measure described in this report is the first one to be provided for census sectors for the whole of Brazil.

BrazDep is a measure of relative deprivation, placing the census sectors on a scale of material well-being from the least to the most deprived. It is useful in comparing areas within Brazil in 2010, but cannot be used to make comparisons across countries or time. Categorical versions of the measure are also provided, placing census sectors into groups of similar levels of deprivation. Deprivation measures, such as the one developed here, have been developed for many countries and are popular tools in public health research for describing the social patterning of health outcomes and supporting the targeting and delivery of services to areas of higher need.

The deprivation measure is exponentially distributed, with a large proportion of areas having a low deprivation score and a smaller number of areas experiencing very high deprivation. There is significant regional variation in deprivation; areas in the North and Northeast of Brazil have on average much higher deprivation compared to the South and Southeast. Deprivation levels in the Central-West region fall between those for the North and South. Differences are also great between urban and rural areas, with the former having lower levels of deprivation compared to the latter.

The measure was validated by comparing it to other similar indices measuring health and social vulnerability at the census sector level in states and municipalities where it was possible, and at the municipal level for across the whole of Brazil. At the municipal level the deprivation measure was also compared to health outcomes. The different validation exercises showed that the developed measure produced expected results and could be considered validated.

As the measure is an estimate of the "true" deprivation in Brazil, uncertainty exists about the exact level of deprivation for all of the areas. For the majority of census sectors the uncertainty is small enough that we can reliably place the area into a deprivation category. However, for some areas uncertainty is very high and the provided estimate is unreliable. These considerations should always be kept in mind when using the BrazDep measure in research or policy. The measure should be used as part of a toolkit, rather than a single basis for decision-making.

The data together with documentation is available from the University of Glasgow http://dx.doi.org/10.5525/gla.researchdata.980. The data and this report are distributed under Creative Commons Share-Alike license (CC BY-SA 4.0) and can be freely used by researchers, policy makers or members of public.

2 Acknowledgements and funding

Many people have assisted with delivering this research project. Doctoral students and researchers at the Centre for Data Integration and Health Knowledge (CIDACS) in Salvador, Brazil, gave us valuable support with country specific knowledge, terminology and translation of the census documents. We would particularly like to thank Andrea Ferreira, Camila Teixeira, Flavia Jose Alves and Poliana Rebouças for their input. Crawford Neilson, Andrew Jackson, Dr. Nema Dean and other colleagues at the University of Glasgow provided advice on the statistical methods and contributed greatly to solving issues with data management. The initial results from this project were presented at different research workshops and meetings, and we are grateful for the feedback and recommendations from the participants.

This research was funded by the National Institute for Health Research (NIHR) (GHRG /16/137/99) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care. The Social and Public Health Sciences Unit is core funded by the Medical Research Council (MC_UU_12017/13) and the Scottish Government Chief Scientist Office (SPHSU13). CIDACS is supported by grants from MCTI/CNPq/MS/SCTIE/Decit/Bill & Melinda Gates Foundation's Grand Challenges Brazil (401739/2015-5), Wellcome Trust, UK (202912 / Z / 16 / Z) and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB INT0001/2015).

3 Introduction

Deprivation measures aim to provide a numeric representation of the complex and multidimensional concept of deprivation – the observed disadvantage of people relative to the society as a whole (Townsend 1987). They are valuable tools highlighting the spatial patterning of material well-being and socioeconomic inequalities in health. Developing small-area composite indices for deprivation and other similar concepts, such as vulnerability, poverty or development has become increasingly popular. First developed in the UK in the 1980s (Carstairs and Morris 1991; Townsend, Phillimore, and Beattie 1988; Jarman 1983), small-area deprivation measures are now used in New Zealand (Salmond and Crampton 2012), Australia (Pink 2013), France (Havard et al. 2008), Japan (Fukuda, Nakamura, and Takano 2007), Spain (Sánchez-Cantalejo, Ocana-Riola, and Fernández-Ajuria 2008) and elsewhere. Increasingly, these measures have also become popular in low and middle income countries, such as Chile (Vasquez 2016), Ecuador (Cabrera-Barona et al. 2015) and South Africa (Noble et al. 2009).

In Brazil, small-area measures of social and health vulnerability have been developed regionally, for a specific state or a municipality. Examples include the Health Vulnerability Index (*Índice de Vulnerabilidade da Saude*, IVS) created for Belo Horizonte (Prefeitura Belo Horizonte 2013) and the Social Vulnerability Index (*Índice Paulista de Vulnerabilidade Social*, IPVS) of the state of São Paulo (State of São Paulo 2013). In addition, individual researchers have developed similar measures for research projects in specific municipalities or towns (Bonfim et al. 2009). However, there are currently no indicators of material disadvantage, deprivation, vulnerability or of a similar concept available nationally for the whole of Brazil at a small-area level (e.g. below the level of municipality) (Ichihara et al. 2018).

Measures for deprivation or disadvantage that cover the whole of Brazil are currently available at the municipal level, with the population generally in tens of thousands. Examples include the social vulnerability index (*Índice de Vulnerabilidade Social*, MIVS) developed by the Brazilian Institute of Applied Economy Research (IPEA) (IPEA 2019) and the Municipal Human Development Index (MHDI) provided by the Atlas of Human Development in Brazil (The Atlas 2019a), a collaboration including the United Nations Development Programme (UNDP), IPEA and the João Pinheiro Foundation (FJP). However, municipalities can be very large (e.g. São Paulo has approximately 12 million people) and therefore heterogeneous, making area-level analysis more difficult. Increasingly, for the large metropolitan areas these measures have also become available for geographies smaller than the municipality, such as Human Development Units (UDH) (The Atlas 2019a).

Our aim is to develop a small-area measure of material deprivation for the whole of Brazil. We do this at the smallest possible geographic area level using census sectors (average population around 600), and a single source of data – the 2010 Brazilian Population Census – for all areas across Brazil. This will increase the extent to which the areas are socio-economically homogeneous and the measure is comparable for all of Brazil. We focus only on measuring *material deprivation*, including indicators of income, literacy and housing characteristics. We do not include indicators of ethnic or demographic composition of an area or indicators that might capture other types of disadvantage. The measure is one of relative deprivation, allowing for comparison of areas within Brazil.

In the future, the Brazilian Deprivation measure, BrazDep, can be used to estimate socioeconomic inequalities in health and mortality, assess the progress of achieving the Sustainable Development Goals across deprivation categories and monitor the health of social welfare program beneficiaries (such as for *Bolsa Família* and *Minha Casa, Minha Vida*). The results from these types of analysis can be used as a benchmark for future inequalities research in Brazil, tracking change over time and across regions with a standardized approach, as well as for cross-national comparisons of inequalities in health and mortality. In addition, BrazDep has the potential to be used as a stratification variable in national surveys, or be included in weights to ensure surveys are representative in terms of deprivation as well as socio-demographic characteristics.

3.1 What deprivation measures can and cannot do

Deprivation measures are frequently used in public health research to describe the social patterning of health outcomes (Leyland et al. 2007; Brown et al. 2014) and support targeting and delivery of services to areas of higher need (Jarman 1983). The measures are particularly useful when individual level measures of socio-economic status (SES) are not available at all or missing for certain population groups (e.g. children, retired people or those not in employment). Combined with other information and tools, deprivation measures can be used as an aid to identify areas for targeted policy interventions. An example of this is Keep Well, a Scottish Government programme, that used the Scottish Index of Multiple Deprivation (SIMD) (The Scottish Government 2012) to identify and target individuals in areas of high deprivation (NHS Health Scotland 2019).

While small-area measures of deprivation are useful tools for both researchers and policy makers, it is important to be mindful of their limitations. Relying solely on a single deprivation measure to decide which areas will or will not receive further funding or support is not recommended. This is because measures of deprivation are estimates of true deprivation and always carry some uncertainty about them. For areas that have very similar deprivation scores, it is not possible to say with certainty which is the most deprived. This is the case also for categorical measures, such as those that identify a certain percentage of most deprived areas. It particularly affects areas that fall close to the cut-off points for different categories. By restricting funding or support simply by using a categorical deprivation measure, many areas that fall just short of the defined cut-off point may be wrongly denied support.

Deprivation measures have also been criticized for their potential biases, such as inability to capture deprivation for rural areas (Martin et al. 2000) or for some ethnic groups (Krieger et al. 2003). For this reason it is important to keep in mind the different populations groups when developing measures of deprivation. We have tried to select indicators that reflect deprivation across Brazil, but we recognize that as Brazil is a very large and diverse country, our measure will have limitations. We recommend that this indicator is used as part of a toolkit, rather than a single tool, for policy interventions.

3.2 About this report

This report presents the BrazDep deprivation measure – the first small-area measure developed for census sectors uniformly for the whole of Brazil. In section 4 it first gives an overview of the 2010 Brazilian Population census, the data source for our deprivation measure. We describe the information collected by the census, the relevant geographies and their sizes, the technical information on data collection, household types and definition, and statistical disclosure control and data availability. It also provides background information for the analysis that follows, such as regional and urban-rural divisions within Brazil.

Following this, section 5 provides an overview of the three deprivation domains included in the final measure (income, literacy and housing conditions). It discusses the specific variables available for analysis and their relationship to each other. In section 6 we move on to combining the selected indicators into a deprivation measure and provide the first results on small-area deprivation in Brazil. In all of our analyses we aim to provide as many results as possible by the five regions of Brazil and by urban-rural classification. Additional results for these and other sections are provided in Appendix A.

In section 7 we validate the measure by comparing it to two other census sector indicators in Belo Horizonte and São Paulo and to municipal level measures of vulnerability and development. We also explore the association between BrazDep and life expectancy, infant mortality and survival at the municipal level.

Issues of uncertainty in measurement are discussed in section 8. We use two different methods of estimating confidence intervals (based on random weights and random numerators) and highlight areas with high level uncertainty about the level of deprivation. This will support the users of our deprivation measure in making informed decisions about applying it in research and policy making. Finally, in section 9 we briefly discuss how the measure associates with ethnicity at the small area level, but given the aggregate nature of the data, we were not able to fully validate the deprivation measure for the different ethnic groups.

4 The 2010 Brazilian Census

The 2010 Brazilian Population Census, conducted by the Brazilian Institute of Geography and Statistics (IBGE), covered approximately 67 million households and 190 million people. Since 1960, the Census has used two questionnaires (IBGE 2010):

- A short basic questionnaire, used in all households (and completed for each of its residents);
- A long sample questionnaire, used in a sample of households (and completed for each of its residents).

The sample questionnaire records detailed information about households which are used to provide estimates for larger geographic areas and the whole population. The questions included in the short basic questionnaire will provide uniform information about the whole population of Brazil. The basic questionnaire covers the following themes:

- Housing characteristics, including access to electricity, water, toilets and bathrooms, means of garbage and sewage disposal.
- Residents' characteristics, including age, gender, ethnicity, kinship relationships, and literacy (for ages 5 or over) of the household members.
- Income for all people aged 10 or over. This information is then used to provide income for heads of households and calculate *per capita* income for the whole household.
- Neighbourhood characteristics for urban areas. These include information on street lighting, paving, sidewalks, etc.
- Household type, such as the number of people in collective residence and private households.
- Mortality and international migration in the last year. (However, these data are not available at the census sector level.)

For more information on how the census was collected see the census technical report (IBGE 2010).

4.1 Census geographies

The census defines 10 different geographical area levels, of which the largest two are the five regions of Brazil and its 27 states, and the smallest the 310,120 census sectors (IBGE 2011). Census data are provided separately for the regions, states, municipalities, metropolitan regions and census sectors. Data from the census sectors can be aggregated to the other geographical areas.

Census sectors (CS), the smallest geographic areas, have been chosen as the units of analysis for developing a small-area deprivation measure for Brazil. The sectors are contiguous areas, created for the purpose of data collection, that respect political and administrative boundaries and the legal framework (IBGE 2011). The number of residents and the territory of a CS are defined in a way that allows the information to be collected by one census taker. It takes into account physical barriers and street layout, and is similar to the routes taken by electric meter readers, postmen or garbage collectors. Census sectors tends to be more homogeneous in areas with greater population density and less homogeneous in areas of low density (The Atlas 2019b). The average population size of the CS is 615.1 (sd = 354.3), making them fairly similar in size to the small-areas used for deprivation measures in many other countries (e.g. SIMD in Scotland). However, the variation in CS population size is quite large, with the largest sector including 5315 people and the smallest only one person.

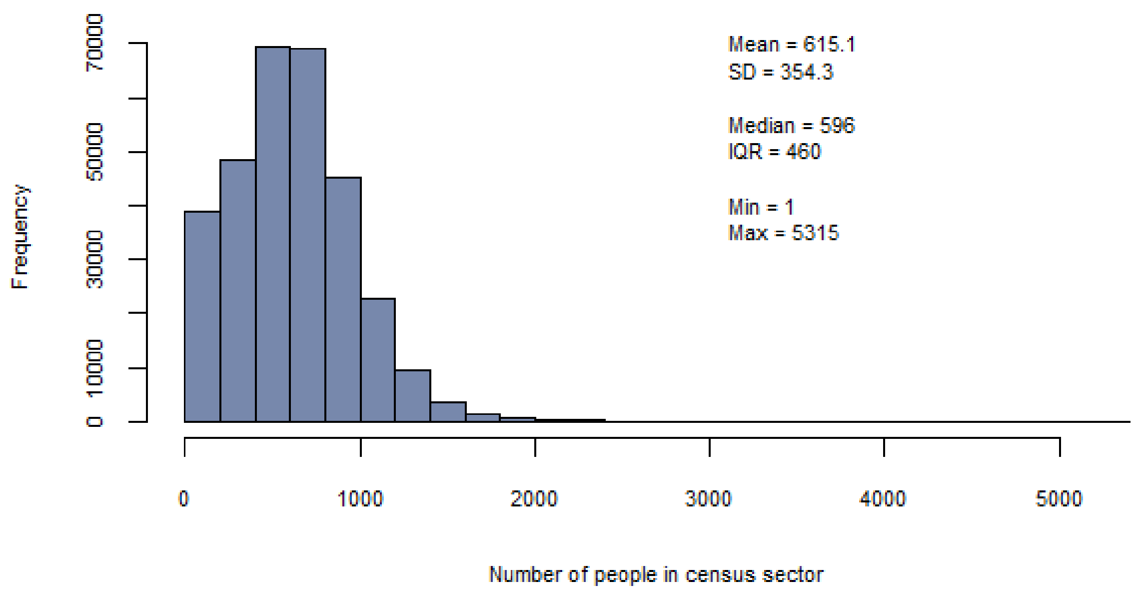


Figure 1: Histogram of census sector population size

Only information collected in the basic questionnaire is available at the census sector level and this excludes information on mortality and migration. More detailed data from the sample questionnaire is available at the municipal level. In 2010 there were 5565 **municipalities** in Brazil, with an average population size of 34,103 (sd = 202,406). Their large population size makes them socio-economically very heterogeneous and, as such, far from ideal for the study of variation in deprivation or socio-economic inequalities in health.

The creation of the **human development units** (*Unidades de Desenvolvimento Humano*, UDH) is an attempt to improve small-area measurement for 16 large metropolitan areas. The UDHs were developed by the Atlas of Human Development in Brazil (The Atlas 2019b) and are not part of census geographies. The goal was to generate socioeconomically homogeneous areas with a minimum of 400 permanent private residents. In 2010 there were almost 10,000 UDH in total, with an average population size of 7137.7 people (sd = 1694.6).[1] Where

[1]The Atlas does not provide data (e.g. number and average population) on the UDHs for the whole of Brazil. The results presented here were collated and calculated by the authors of this report based on metropolitan area information published by the Atlas.

Table 1: Population size, and the number and type of sectors by region

Region	Population size				Census sectors		Urban areas	
	Mean	SD	Total	%	Number	%	Number	%
Northeast	660.7	373.7	53081950	27.8	80348	25.9	49212	61.2
North	742.5	446.5	15864454	8.3	21365	6.9	13175	61.7
Central-West	587.7	331.4	14058094	7.4	23919	7.7	18806	78.6
Southeast	605.1	327.7	80364410	42.1	132809	42.8	116576	87.8
South	529.9	330.2	27386891	14.4	51679	16.7	38945	75.4
Total Brazil	615.1	354.3	190,755,799	100	310,120	100	236,714	76.3

possible, a UDH is created as an aggregation of census sectors. To ensure the UDH are homogeneous, they are not always spatially contiguous.

The larger size of the UDH compared to the census sectors allows data from the sample questionnaires to be used. Indices released at the UDH level can include a more varied set of indicators, such as school attendance and level of education, as compared to literacy, the only education indicator available from the basic questionnaire. To develop UDH level indicators, research organizations have to submit a proposal to IBGE and, if approved, will be allowed access to the micro data at an IBGE safe haven. New indices can then be developed by the researchers and will be released to the public after undergoing statistical disclosure control by IBGE.

Our goal was to measure deprivation for the whole of Brazil at the smallest possible geography, and for this reason we chose to develop the measure at the census sector level. It limits the number of indicators available, but increases homogeneity of the area and ensures consistent coverage of data across all of Brazil. To compare the BrazDep deprivation measure to other similar indicators, such as MHDI and IVS, we also calculated it for municipalities.

Because of great regional differences within Brazil, our analysis will compare deprivation indicators and measures between regions. Table 1 provides the distribution of census sectors and population across Brazil. Over half of the population lives in the two Southern regions, with 42% in the biggest Southeast region. Just over a third of people live in the two Northern regions, with 27.8% in Northeast, the second biggest region in Brazil. The Centre West is the smallest region of Brazil (7.4% of population), including the Federal District and the Federal Capital Brasília. The Southern regions of Brazil are more developed and urban compared to the North. The Centre West region share similarities with both South and the North.

4.2 Urban and rural areas

The Brazilian census identifies three types of urban and five types of rural areas, listed together with the distribution of census sectors and population in Table 2. The vast majority of people (82%) live in urbanized cities or towns and just 13% in rural areas. Because such a large majority of areas and people fall into either of the two dominant urban and rural area types, we have decided to use a binary definition that groups all urban areas into one and all rural areas into another category.

The census urban rural classification is not based on population size, density or distance

Table 2: Distribution of census sectors and population by urban rural classification

Area type	Binary type	Number CS	Percent CS	No People	Percent people
Urbanized area in a city/town	Urban	227250	73.28	157191046	82.40
Non-urbanized area in a city/town	Urban	6191	2.00	2435872	1.28
Isolated urban area	Urban	3273	1.06	1298886	0.68
Rural cluster in an urban expanse	Rural	1514	0.49	869377	0.46
Isolated rural cluster, village	Rural	9200	2.97	4173565	2.19
Isolated rural cluster, core	Rural	236	0.08	92149	0.05
Isolated rural cluster, other	Rural	1281	0.41	308857	0.16
Rural area, excluding cluster	Rural	61175	19.73	24386047	12.78
Total		310,120	100	190,755,799	100

to services or major centres, but on administrative law dating back to 1938 (Decreto-Lei No. 311, de 02.03. 1938). For this reason the census classification may not actually reflect population size, density or other aspects generally related to urban rural classification. A new definition of urban rural classification based on accessibility and population density will be introduced for the 2020 census alongside the current legal definition.

4.3 Census household and sector types

The census collects information on all individuals regardless of whether they live in purpose built homes, hospitals, prisons, dormitories, other institutions, or in makeshift and improvised shelters (not purpose built for living). Based on the place of residence, the census distinguishes three distinct types of households (IBGE 2011):

- Private household (PHH) – households where the relationship between occupants is dictated by ties of kinship, domestic dependency or norms of coexistence.
- Permanent private households (PPHH) – a subcategory of PHH, distinguishing households residing in homes that were built to serve, exclusively, as a dwelling and serve as a dwelling to one or more people. It is by far the most dominant type of households, with 99.5% of the total population living in PPHH.
- Collective residence (CR) – people/households residing in institutions where the relationship between people, whether residents or not, is restricted to administrative subordination rules. It includes hotels, motels, campsites, prisons, barracks, military posts, asylums, orphanages, convents, hospitals and clinics, housing for workers or students, and so forth. Approximately 0.27% of people live in collective residences.

A small minority of people live in private households that are not purpose built as a dwelling. These households are referred to by the census as improvised households (IHH). These are private households, but not *permanent* private households. The census provides a count of improvised households for all sectors and across Brazil 102,874 households (0.18%) are recorded as improvised. However, the census does not give a count of individuals who live in improvised households. We have estimated this by taking the difference between the number of people in private households and those in permanent private households.

Many of the census data, such as population distribution by age and sex, are provided for all people across all types of households. Because of this we aimed to base the deprivation

Table 3: Distribution of census sectors and population by census sector type

	All Brazil				Omitted due to SDC			
	Census sectors		Population		Census sectors		Population	
Sector type	Number	%	Number	%	Number	%	Number	%
Common or non-special sector	290192	93.57	178158683	93.40	5834	2.01	350236	0.20
Special agglomerate	15816	5.10	11431619	5.99	63	0.40	15182	0.13
Barracks, military bases, etc.	58	0.02	15541	0.01	6	10.34	255	1.64
Special accommodation, camps, etc.	49	0.02	8820	0.00	17	34.69	609	6.90
Boats, ships, etc.	3	0.00	128	0.00	3	100	128	100
Indigenous village	1211	0.39	223267	0.12	123	10.16	3153	1.41
Prisons, etc.	315	0.10	205755	0.11	82	26.03	64594	31.39
Orphanages, convents, hospitals, etc.	475	0.15	47528	0.02	115	24.21	9850	20.72
Rural settlements projects	2001	0.65	664458	0.35	59	2.95	3345	0.50
Total	310,120	100	190,755,799	100	6302	2.03	447,352	0.23

measure on as many people as possible. However, some information, such as household income and housing conditions are only provided for private or permanent private households and not for collective residences. This means that people in collective residences are excluded from some analysis and ultimately it was not possible to provide a deprivation measure for the small number of sectors that only included communal residences.

In addition to distinguishing between households, the census also distinguishes between nine different types of census sectors. Table 3 provides the distribution census sectors and population by census sector type. These types overlap to an extent with household types in that they capture the (dominant) type of household in the sector. For example, most of the census sectors that are classified as orphanages only contain people in communal residences.

The common or non-special sector type is by far the most dominant type with 93% of all people living in these areas. The second most common type is special agglomerates or favelas, including 5.9% of people. Other sector types are extremely rare with fewer than 0.5% of people.

4.4 Statistical disclosure control

The size of the census sectors varies considerably, with some areas including very few people. To protect the privacy of these individuals, statistical disclosure control (SDC) is applied and data for census sectors with fewer than five permanent private households are omitted. This affects 6302 (2%) census sectors and approximately 0.2% of the total population surveyed by the census. Table 3 shows the number and percentage of census sectors and people omitted from the results due to SDC by census sector type. SDC is more likely to affect special types of census sectors, such as boats, prisons or orphanages. Areas affected by SDC are not included in the deprivation measure.

In addition to SDC, 600 further census sectors had to be removed from the analysis as they only included people living in communal residences and no private households. As a result, a deprivation measure is calculated for a total of 303,218 sectors (97.8%), covering 190,145,077 people (99.7%).

4.5 Data access and quality

The census data were downloaded together with relevant documentation from the IBGE website.[2] The data are provided in 26 thematic tables separately for each state. (See IBGE (2011) for a description of tables and variables.) They include a unique census sector identifier that allows linkage across the thematic tables. To facilitate analysis, a MySQL database was built for the whole of Brazil.

A number of data quality issues were present in the original census data. Most often this meant some cell values were corrupt (e.g. text values instead of numbers) or data were missing for some census sectors. In two cases variables were missing from the tables. For table *Entorno05* on neighbourhood characteristics some variables were missing for the state of Goiás. For table *Domicilio02* on household characteristics the census sector identifier that allows linkage across different thematic tables was absent for São Paulo.

In most cases we did not attempt to extrapolate or impute the missing data. However, in the case of São Paulo it was necessary to add the missing census sector identifiers. In this case we simply made the assumption that census sectors in table *Domicilio02* are in the same order as for all other tables for São Paulo and copied the census sector identifier from another table. We checked to see if the assumption held by merging the *Domicilio02* table to other tables and comparing the variables for urban-rural classification and for population counts present in *Domicilio02* to the same variables in other tables. No discrepancies were noted.

Due to the noted quality issues, a number of checks were always ran on the data prior to analysis to confirm internal consistency. These checks included calculating the totals from disaggregated data (e.g. from age and gender distributions) and comparing these to the totals provided in the different tables. Variables that were present in multiple thematic tables (mostly urban-rural classification, and population and household counts, but in a few cases also total and average income) were compared across tables. We were able to confirm internal consistency for all data used in the deprivation measure and in the analysis below. However, we noticed issues in table *Pessoa04* on literacy rates for women by ethnicity and age, and these data were not used in the report.

[2] ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/

5 Deprivation domains and indicators

There is a lot of variation internationally in the specific indicators included in deprivation measures, but nearly all include a set of common dimensions of deprivation, often called domains. These domains cover (un-)employment, material wealth (e.g. car ownership, income), indicators of socioeconomic position (SEP), such as education and/or occupation, and housing conditions, such as overcrowding, home ownership or renting from a public authority. The Brazilian census basic questionnaire covers three of these themes – wealth approximated by income, SEP and education indicated by literacy rates, and housing conditions captured by access to different amenities (e.g. toilets and shower/bath) and services (garbage collection, access to water and sewage). The census further includes information on neighbourhood characteristics (e.g. presence of street lighting, sidewalks and road paving/surfacing), that also closely relate to the concept of material deprivation.

For each of these domains, we considered a number of different variables and definitions available. The main criteria for selecting between indicators to reflect each of these domains was guided by our research goals and previous literature in the field (Noble et al. 2006; Allik et al. 2016). These include:

- **Conceptual fit.** The indicators should, as far as possible, capture deprivation for both men and women, for different regions, population and ethnic groups.
- **Wide coverage.** Our aim was to calculate a deprivation measure for as many census sectors as possible and base each of the indicators on as many people within a sector as possible. For this reason we prefer indicators that include as many people as possible.
- **Empirical prevalence and variation.** The indicator should capture a major feature of deprivation, experienced by a number of people across Brazil, and it should have sufficient variation to distinguish between areas.
- **Correlation with other indicators.** While each of the domains measure a different aspect of material deprivation, they should all still be highly correlated as they try to capture the same broad concept. We prefer indicators that have a higher correlation with other domains as this can be considered an indication of closeness to the same theoretical concept.

We often had to choose between these criteria, as the different variables and definitions did not each meet all the desirable qualities. For example, while personal income has wider coverage compared to household and head of household income, it has a weaker conceptual fit to deprivation due to the nature of how the data were collected. In addition, neighbourhood characteristics were unfortunately only recorded for urban areas and had to be excluded from the final measure.

5.1 Income

The census provides three data tables on income – *DomicilioRenda*, *ResponsavelRenda* and *PessoaRenda* – covering respectively household, head of household and individual income. All tables give the total income for a census sector and also the number of households, heads of households and people by income categories relative to the minimum wage (510 Brazilian Reais in 2010, approximately 290 USD at the time). Total income for census sectors cannot be used as a deprivation indicator on its own as census sectors vary in size. Total income may be used to calculate average (mean) income for a sector, but the mean can easily be increased by a few very high values and may mask deprivation. For this reason, we have used income categories relative to the minimum wage for the deprivation measure, giving the number of household, head of households or people in a census sector who might be consider deprived.

5.1.1 Household income

Household income is provided for private households, that is for both permanent private and improvised households, but not for collective residences. Thus, sectors with only collective residences (600 in total) have no household income data, limiting the coverage. The variables give the number of households by income groups defined by *per capita* income relative to the minimum wage (MW). The household income group variables do not distinguish between permanent private and improvised households. The cut-off points for the income groups are: nominal monthly household *per capita* income up to 1/8 minimum wage, income over 1/8 and up to 1/4 MW, over 1/4 and up to 1/2 MW, over 1/2 up to 1 MW, over 1 to 2 MW, over 2 to 3 MW, over 3 to 5 MW, over 5 to 10 MW, and over 10 times the MW. A separate variable provides the number of households with no income.

Table 4: Percent of households by per capita household income relative to MW, by region

	No income	≤ 1/8	1/8-1/4	1/4-1/2	1/2-1	1-2	2-3	3-5	5-10	10+
Percent	4.3	3.7	6.8	17.2	28.7	21.9	7.0	5.3	3.6	1.5
Cumulative percent	4.3	8.0	14.7	31.9	60.6	82.5	89.5	94.9	98.5	100
Region	Cumulative percent									
Northeast	5.7	15.4	28.3	53.1	80.3	91.7	95.0	97.6	99.3	100
North	6.5	13.5	25.8	48.9	75.2	89.7	94.1	97.3	99.3	100
Central West	3.5	4.6	9.1	25.2	56.1	79.7	87.2	93.2	97.7	100
Southeast	4.0	4.9	8.8	22.7	51.9	78.0	86.7	93.3	98.0	100
South	2.3	3.2	6.4	18.2	47.7	77.9	87.6	94.5	98.5	100

Table 4 shows the percent and the cumulative percent of households across Brazil by the income groups defined in the census. About a third (32%) of all households have per capita income below 1/2 of minimum wage and 61% have *per capita* income below MW. This varies considerably across regions. While in the two Northern regions about half the household have a *per capita* income below 1/2 MW, in the Southern regions and in Central West this is just 18-25%. To capture deprivation, we may consider using the number of household with *per capita* income below 1/4 of minimum wage (15% across Brazil) or the number of households with *per capita* income below 1/2 of minimum wage. Considering all households below the minimum wage might include too many households in the deprived group, particularly in the Northern Brazil.

The histograms in Figure 2 show the distributions of the two potential income deprivation indicators by region. It is evident that low household income is much more common in the Northern regions compared to the Southern regions. It also shows that using 1/4 MW as a cut-off reduces variation in the Southern regions and a large proportion of sectors will have very low income deprivation. The cut-off point 1/2 MW may work better as a deprivation indicator as it is better able to distinguish between sectors in South and Southeast.

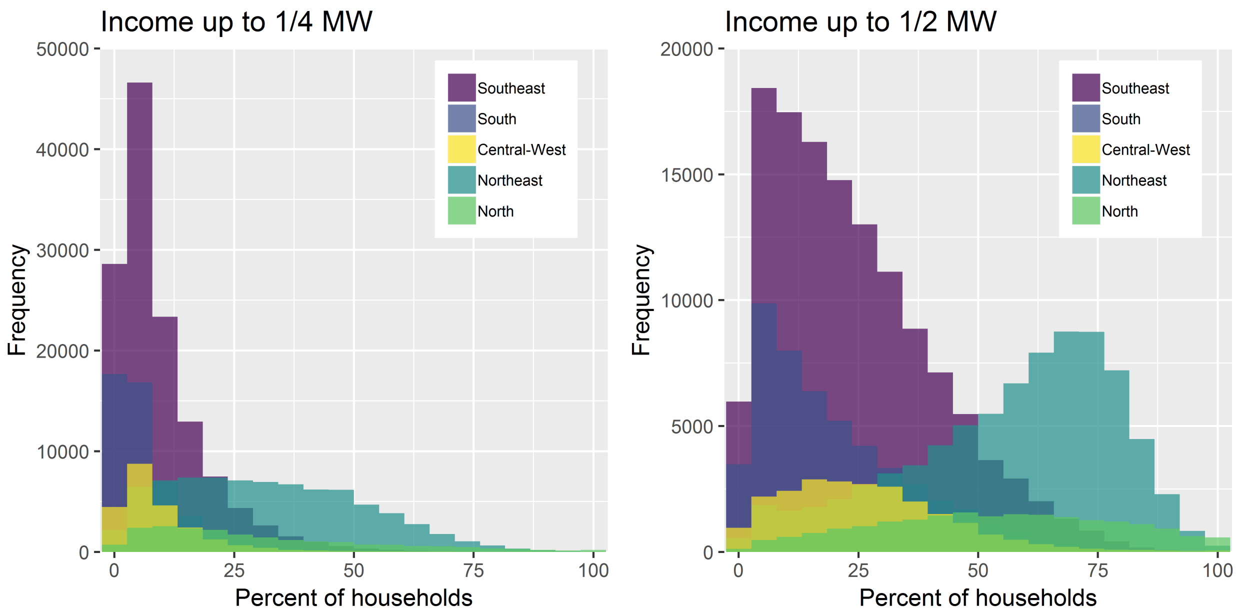


Figure 2: Histograms of household income, up to 1/4 and 1/2 MW, by region

Across Brazil, the two household income variables are strongly correlated at 0.92 (see Table A.1 in Appendix A), but the correlation is weaker in the Southern and Central West regions of Brazil (r = 0.86-0.87), and stronger in the Northern regions (r = 0.91-0.92). Despite the generally strong correlations, for a few areas the difference between the measures will lead to a considerable change in rank.

5.1.2 Head of household income

Data about the head of household (HoH) income are recorded for areas that have any heads of private households and are not collected for collective residences. For this reason 492 census sectors with no heads of household are excluded. Unlike for household income, all information is provided by household type, separately for heads of permanent private and improvised households, and also by the head of household sex. Most of the income groups given are the same as for *per capita* household income, but the lowest income group is below 1/2 of MW and highest 20 times the MW or above.

Table 5 shows the percent and cumulative percent of head of households by income categories. Nearly half (45%) of head of households across Brazil have an income below the MW and 19% receive an income below 1/2 MW. The results by region are very similar to household income – low HoH income is much more common in the North and Northeast compared to the South, Southeast and Central West regions of Brazil. For example, in the Northeast 68% of HoH receive an income below the MW, but in the South this is only 32%.

Based on the cumulative percentages in Table 5 we identified as HoH below 1/2 and 1 MW as potential cut-off points to indicate deprivation. Figure 3 shows the distributions of the

Table 5: Percent of head of households by income relative to MW, by six regions

	No income	$\leq 1/2$	1/2-1	1-2	2-3	3-5	5-10	10-15	15-20	20+
Percent	12.8	6.0	26.7	25.4	10.3	8.7	6.6	1.3	1.1	1.0
Cumulative percent	12.8	18.9	45.6	71.0	81.3	89.9	96.6	97.9	99.0	100.0
Male heads of household										
Percent	8.9	4.4	24.3	27.6	12.4	10.3	7.8	1.6	1.4	1.3
Cumulative percent	8.9	13.3	37.6	65.2	77.6	87.9	95.7	97.3	98.7	100.0
Female heads of household										
Percent	19.1	8.6	30.5	22.0	6.9	6.1	4.8	0.9	0.6	0.5
Cumulative percent	19.1	27.7	58.2	80.2	87.1	93.2	98.0	98.9	99.5	100.0
Region	Cumulative percent for all HoH									
Northeast	15.5	29.7	68.3	85.3	90.3	94.7	98.2	98.9	99.5	100.0
North	17.0	26.5	58.4	80.0	87.2	93.3	98.0	98.8	99.5	100.0
Central-West	11.2	14.1	38.7	66.8	77.8	86.8	94.9	96.8	98.4	100.0
Southeast	12.3	14.6	35.9	64.4	76.9	87.6	95.7	97.3	98.8	100.0
South	8.9	11.7	32.4	63.6	77.5	88.6	96.5	98.0	99.1	100.0

two head of household income variables by regions. Both include HoH with no income. Distribution for HoH income below 1/2 MW is very skewed in the South and Southeast, with most areas having very few HoH in that wage category. Again, using the higher cut-off point of 1 MW increases variation in the South and may, for this reason, be a slightly better deprivation indicator compared to the cut-off point 1/2 MW.

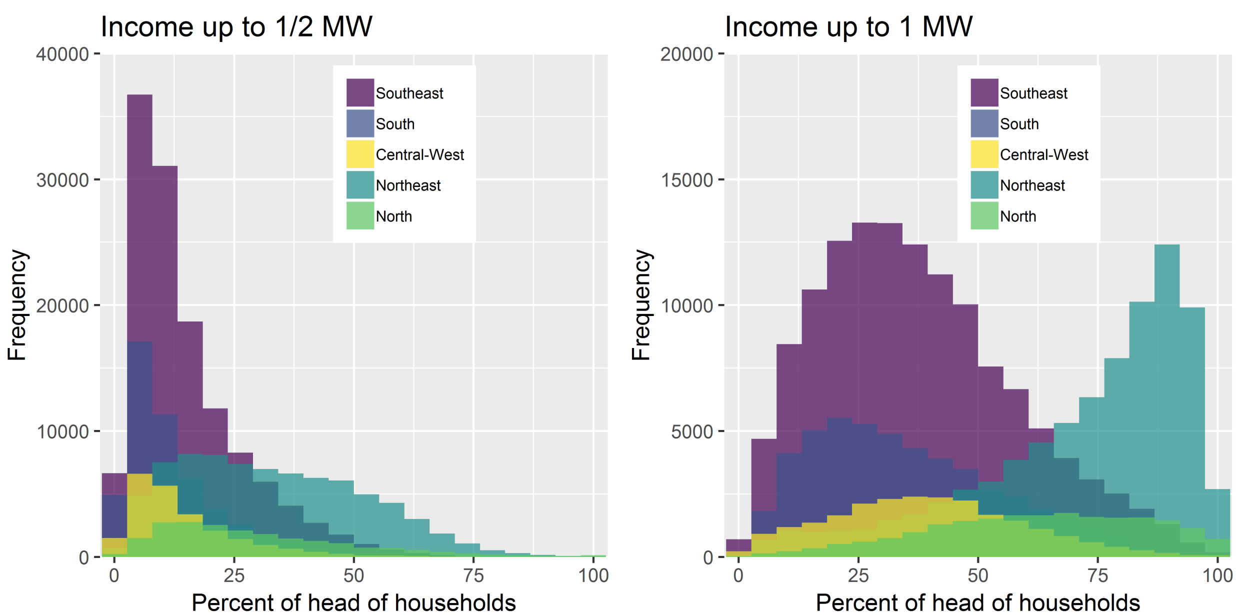


Figure 3: Histograms of head of household income, up to 1/2 and 1 MW, by region

Correlation between the two variables is again strong (r = 0.8, see table A.1 in Appendix A), but weaker within regions.

Table 5 also shows that low income is much less common among male HoH compared to females. Over half (58%) of the female head of households have an income below the minimum wage, but for men this is 46%. Some researchers in Brazil have used gender specific indicators (e.g. income for female HoH or literacy among women) to measure socio-economic disadvantage (Santos et al. 2007). We considered using income for only female heads of household as a deprivation indicator, but since some areas (2128 sectors) have no and others very few female heads of household[3], this idea was not pursued.

5.1.3 Personal income

Personal income is recorded for all people aged 10 and above, regardless whether they reside in a private household or in a collective residence. This means that personal income has the widest coverage of all income variables and is recorded for 303,815 census sectors. Data is also provided separately for people in permanent private households and by gender, but not by age groups. The personal income cut-off points are the same as for head of household income, the lowest group being below 1/2 of minimum wage and highest 20 times the minimum wage or above.

Table 6 shows the percent and cumulative of people by income levels. Over a third (37.1%) of all people aged 10 or above have no income and nearly two thirds (64.5%) have an income below the minimum wage. For men, these percentages are slightly lower, with 30.7% having no income and 56% with an income below the minimum wage. For women, the percentages are much higher, respectively 43.1% and 72.6%. As for other income variables, there are considerable regional differences. In the Northern regions 76.4-81.8% have an income below the minimum wage, while in the South and Central West Brazil these percentages are between 52.2-59.8%.

Table 6: Percent of people by personal income relative to MW, by six regions

	No income	≤ 1/2	1/2-1	1-2	2-3	3-5	5-10	10-15	15-20	20+
Percent	37.1	6.3	21.1	18.9	6.3	5.0	3.6	0.7	0.5	0.4
Cumulative percent	37.1	43.4	64.5	83.4	89.7	94.7	98.3	99.0	99.6	100
Men										
Percent	30.7	4.5	20.8	22.1	8.4	6.5	4.6	0.9	0.8	0.7
Cumulative percent	30.7	35.3	56.0	78.1	86.5	93.0	97.6	98.5	99.3	100
Women										
Percent	43.1	8.0	21.5	15.9	4.3	3.7	2.6	0.4	0.3	0.2
Cumulative percent	43.1	51.1	72.6	88.4	92.8	96.4	99.0	99.5	99.8	100
Region	Cumulative percent for all people									
Northeast	42.3	55.4	81.8	92.2	95.0	97.4	99.2	99.5	99.8	100
North	45.4	54.4	76.4	89.6	93.6	96.8	99.1	99.5	99.8	100
Central-West	34.8	38.7	59.8	80.5	87.3	92.8	97.4	98.4	99.2	100
Southeast	35.1	38.0	56.4	79.0	86.9	93.3	97.8	98.7	99.4	100
South	29.9	33.4	52.2	78.1	87.1	93.9	98.3	99.0	99.6	100

Based on the cumulative percentages we decided to use the groups no income and income below 1/2 MW (including no income) for indicators of deprivation. However, the main limitation of this definition is that the no income group is conceptually ambiguous. It does not distinguish between people who cannot find paid work, and those who do not work for other reasons (e.g. choose not to, are ill) and because of this we cannot be certain that no income means deprivation. The problem is likely to be worse for women, who are probably

[3]The percent of male head of households is above 80% in 15% of census sectors across Brazil. This is more common in rural areas, where the percent of male head of households is above 80% in almost half of the census sectors.

more likely to choose not to work (e.g. for cultural or other reasons). Since personal income is recorded for all people aged 10 and above, many children will also be recorded in the zero income group, regardless of whether they wish or need to work.

Because of this, previous researchers have excluded people with no income (Bonfim et al. 2009). Here, we have combined people with income below 1/2 and 1 MW into a deprivation indicator. This captures the percentage of people with low income. We considered using people with income up to 1/2 MW on its own (excluding those with no income), but as this group is quite small (6.3% of all people) and has less variation, we did not pursue this option. The obvious shortcoming of excluding those with no income is ignoring some very deprived people.

Figure 4 shows the distributions of the three income variables by region. All histograms show differences between regions, but these are less obvious for the percentage of people with no income. This may be due to the ambiguous nature of the variable. The distributions for the other two income variables are fairly similar to each other and to the histograms for household and HoH income.

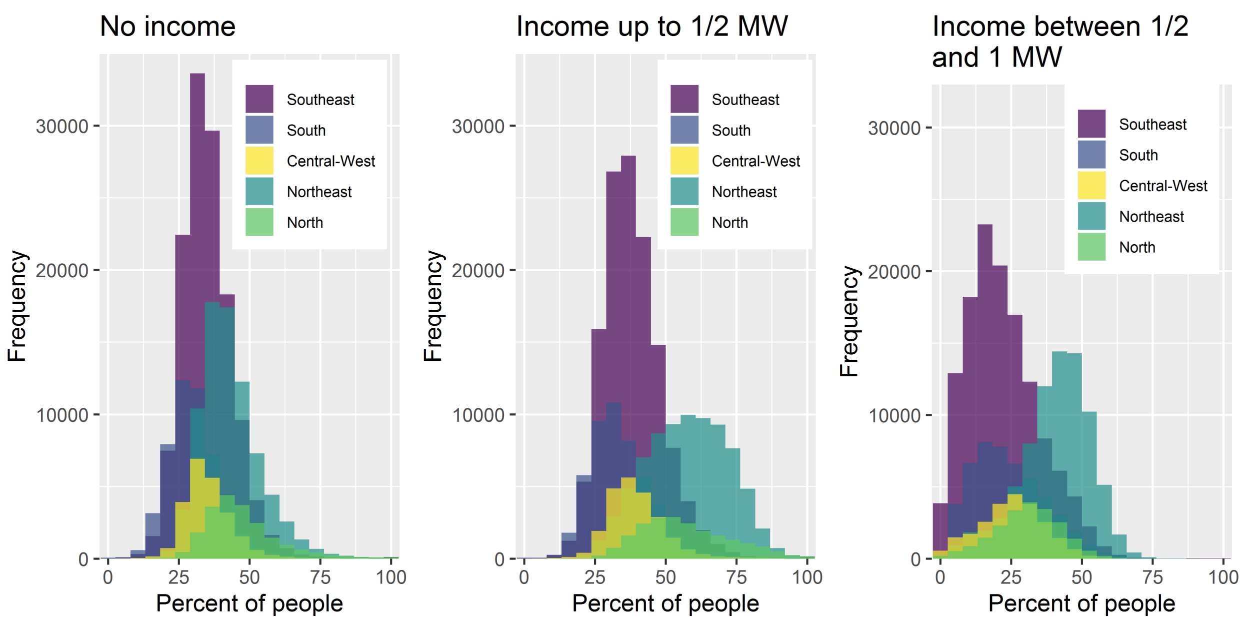


Figure 4: Histograms of personal income, by region

As with head of household income, a gender specific indicator was ruled out as some areas have very few or even no women (or men). As low personal income is fairly common among women, it might not be a good discriminatory indicator of deprivation.

5.1.4 Comparison of income measures

We have considered a number of income indicators with different cut-off points and reflecting household, head of household and personal income. Table A.1 in Appendix A gives the correlations between all the income variables across Brazil.

Using the desirable criteria outlined previously, we can reduce the number of indicators suitable for a deprivation measure. First, given the conceptual issues around personal income, we decided to focus on household or head of household income measures. Completely ignoring people or households with no income is problematic as we are potentially excluding the most deprived. Including no income in the household or head of household income variable is less likely to give a "false positive" for deprivation compared to no personal income. While we may expect many people (e.g. children) not to work and have an income because they do not need to, this is unlikely to be the case for head of households or for all people in the household.

Household and head of household income also have the benefit of reflecting the conditions of a household unit – people generally live in households and the conditions, such as low or no income, are shared and affect everyone in a household. In the final analysis, we preferred household income over head of household income as the former is calculated *per capita*, meaning that it takes account of household size. Head of household income does not account for household size and a high HoH income may not mean a lack of deprivation if the household is also large. We have included both household income variables (with cut-off points below 1/4 and 1/2 MW) in the analysis to follow.

5.2 Literacy

Literacy is covered in the census tables *Pessoa01* and *Pessoa02*. It is recorded for all people aged five and above, and separately by sex, ethnicity and single year age groups. We considered three different definitions of literacy rates as deprivation indicators. The summary statistics for these are shown in Table 7 and the histograms in Figure 5.

The first of these variables is simply the literacy rate for all people aged five or above. Across all census sectors, literacy rates are relatively high at 88.4% and the vast majority of areas have literacy rates above 70%. Secondly, we have also calculated literacy rates for those aged seven and older to exclude children who may have yet to start school. This has minimal effect on the measure, but might be preferred as mandatory school age in Brazil is six.

Regional variation in literacy rates is quite notable. The mean literacy for people aged five and over is around 80% in the North, while in the South this is over 90%.

As these two variables have quite high means and low variation, they may not be effective in distinguishing between deprived and less deprived census sectors. For this reason we considered combining literacy for ages 7-9 and for those aged 50 and above into a single indicator with more variation. The summary statistics and the histogram both show that this variable has more variation across Brazil and within the different regions. The disadvantage of this indicator is, however, that in some census sectors it is based on a very small number of people and in a few cases it cannot be calculated as there are no people in these age groups.

We did not consider ethnicity or gender specific indicators of literacy for the deprivation measure. Gender based measures were excluded here as across all age groups literacy rates are the same for women and men (88.9% and 88.1% respectively). Among younger people and children women have slightly higher literacy; for older people, men tend to have higher

Table 7: Summary statistics for literacy rate variables

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD	Missing
Literacy 5+	0	84.5	92.4	88.4	96.3	100	11.2	6303
Literacy 7+	0	86.3	93.9	89.8	97.3	100	10.9	6304
Literacy 7-9 and 50+	0	68.3	84.5	78.6	93.7	100	19.2	6395
	Literacy 5+							
Northeast	0	68.5	78.6	78.1	89.2	100	13.2	1223
North	0	78.0	86.9	83.4	92.8	100	13.9	437
Central-West	0	87.5	92.2	90.9	95.8	100	7.2	468
Southeast	0	90.9	94.7	93.2	97.3	100	6.2	2401
South	0	91.1	94.6	93.4	97.1	100	5.1	1774

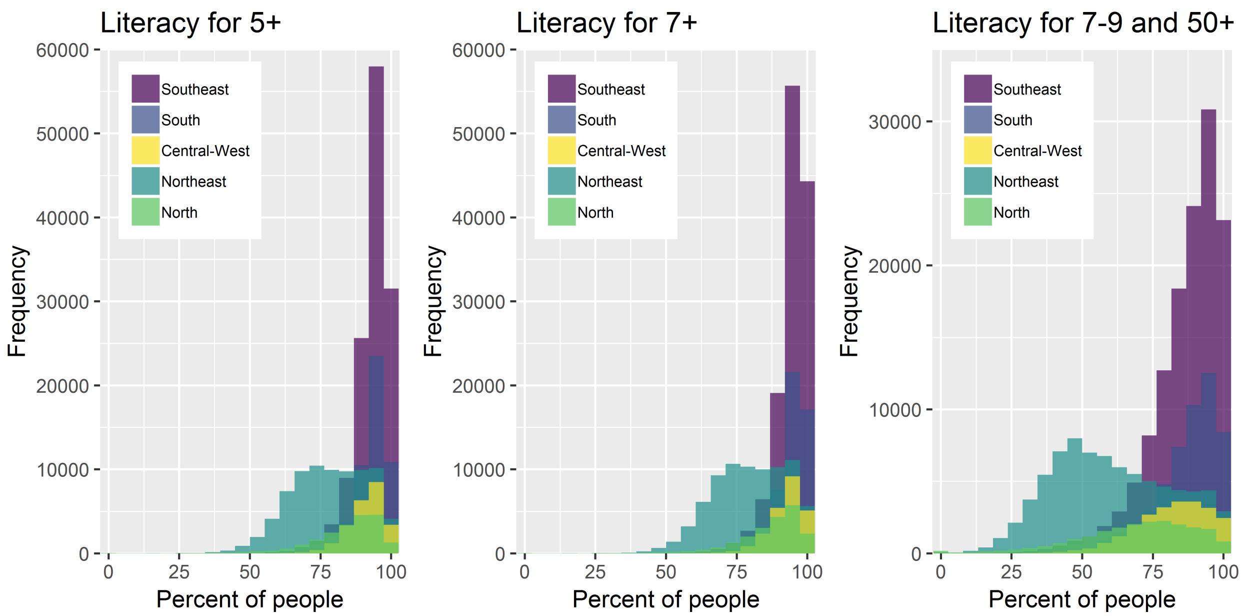


Figure 5: Histograms of literacy rates, by region

literacy. As the gender differences are confined to very specific and often small age groups, they were not sufficient to be included in the measure. Some researchers have used literacy rates among women in deprivation measures (Santos et al. 2007) and this may make sense within specific regions (e.g. in the North) where gender differences in literacy are more pronounced. Ethnicity based measures of literacy were excluded as many sectors only have one or two ethnic groups and therefore there is not sufficient variation in these indicators for them to be useful.

5.3 Housing characteristics

The census includes two tables covering housing characteristics, *Domicilio01* and *Domicilio02*. The former gives the number of households and the latter the number of people by household conditions. The two are strongly correlated. For the deprivation measure we have focused on the table *Domicilio02*, which gives the number of people experiencing different housing characteristics.

Data are provided only for people in permanent private households and not for people in improvised households or in collective residence. We have decided to include people in improvised households in the denominator, thereby assuming that all people in improvised households lack any of the housing characteristics. This is reasonable as those without a permanent purpose built home should be considered housing deprived. It also improves the coverage of the measure by including more people and sectors. People in collective residences have been excluded.

The housing characteristics covered in the census include:

Access to electricity. The census records the number of people with access to electricity and those without it. In addition, it records the number of people who have access to electricity from a distribution company versus any other method. Those with electricity from a distribution company are further split into three groups: people in households with a meter installed for a single household (exclusive meter), people in households with a shared meter between multiple households and those without a meter. After initial analysis, access to electricity and to network electricity were dropped from the analysis as they had too little variation (see Table 8). Other measures relating to access to electricity were deemed not to have a clear relationship to deprivation.

Access to water. The census groups people into four categories depending on whether access to water is via a network, well or spring on the property, collected from rainwater or by other means. We consider two alternative coding methods to capture deprivation. First, we consider all those without access to network water as deprived. However, in rural areas water from a well might not indicate deprivation. For this reason, the second definition consider all those without access to water from either a network or a well as deprived. We have termed the first of these coding methods as the narrow and the second as the extended definition of access to water.

Toilet and/or a bath at home. The census provides information about the number of people with access to both a toilet *and* a bath/shower at home, and the number of people with access to either a toilet *or* a bath/shower at home. The two variables are quite similar in their distribution, but given the relationship they have to sewage disposal (discussed below), we decided to use the number of people with access to a toilet *and* a bath/shower at home.

Means of sewage disposal. The census classifies people and households by six means of sewage disposal: (1) access to network sewage, (2) access to a septic tank, (3) a cesspit, (4) disposal into a ditch, (5) disposal into a river, lake or sea, and (6) other means of disposal. Again we considered two methods of sewage disposal to capture deprivation. The first of these, the narrow definition, included those without access to network sewage as deprived. But in rural areas it may not be reasonable to expect people to have access to network sewage and thus adequate sewage disposal should also include a septic tank. The second, extended definition, then counts all those without access to either network or septic tank as deprived.

Access to network sewage and/or a septic tank is, however, a subset of all people who have access to a toilet *or* bath/shower at home. For this reason, using the percent of people who have both a toilet *and* a bath/shower at home might be preferred as this variable only *partly* overlaps with those who have access to network sewage and/or a septic tank. Using this

Table 8: Summary statistics for housing condition variables

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD	Missing
Access to electricity	0	99.3	100	97.6	100	100	10.0	6794
Access to network electricity	0	98.8	99.9	96.3	100.0	100	14.0	6794
Access to network water	0	60.4	97.9	75.1	100	100	37.7	6794
Access to water from network or well	0	94.1	99.6	89.2	100.0	100	23.7	6794
Toilet or bath/shower	0	99.0	100	95.7	100	100	13.6	6794
Toilet and bath/shower	0	95.5	99.5	91.0	100	100	20.5	6794
Access to network sewage	0	0.3	46.7	48.5	98.3	100	44.6	6794
Access to network sewage or septic tank	0	7.4	86.4	61.1	99.6	100	42.4	6794
Garbage collected	0	83.9	99.6	79.5	100.0	100	36.4	6794
Housing conditions, extended	0	70.3	93.8	80.2	99.7	100	25.5	6794
Housing conditions, narrow	0	55.1	82.4	73.5	99.0	100	29.1	6794
	Housing conditions, extended							
Northeast	0	36.5	73.1	64.5	93.6	100	30.2	1277
North	0	37.0	69.3	59.8	82.7	100	30.0	449
Central-West	0	72.6	78.8	79.2	98.8	100	20.5	496
Southeast	0	92.3	99.4	91.5	100.0	100	16.2	2732
South	0	75.0	94.0	84.9	99.4	100	18.1	1840

combination does not remove all of the double counting of individuals, but it is the optimal solution. Most data for census sectors are provided as stand alone tables, i.e. without cross tabulations. This means any measure using a combination of two or variables for housing characteristics will lead to double counting of some individuals. Therefore, if we use multiple housing conditions, we need to combine the variables in a way that minimizes the effects of double counting.

Means of garbage disposal. Garbage disposal is split into six categories: (1) collected by a service provider, (2) burned, (3) buried, (4) thrown into wasteland or public space, (5) dumped in a river, lake or sea, and (6) other methods of garbage disposal. All means of disposal other than garbage collection by a service provider are classified as indicating deprivation.

Table 8 shows the summary statistics for the different housing characteristics and Figure 6 gives the histograms for the four variables that we intend to include in the deprivation measure. As mentioned, access to electricity is very common and for this reason has been excluded from the measure. Access to a toilet or a bath/shower and access to a toilet and a bath/shower are also both fairly common, with means respectively of 95.7% and 91%. The distribution for the latter is very skewed, with access at about 100% for the vast majority of areas. This is especially the case in the Southern regions and in the North there are also areas where access to a toilet and a bath/shower is close to zero. Access to water from the network and from the network or a well have somewhat lower means and slightly more variation compared to access to toilet and bath/shower, but generally the distribution in the histogram looks the same.

Access to sewage by means of network or septic tank and garbage collection by a cleaning service have the lowest means and highest variations. The histograms for these variables are, however, bimodal – while in the Southern regions most people will have access to these services, in the North this is much less common. As with income and literacy, regional

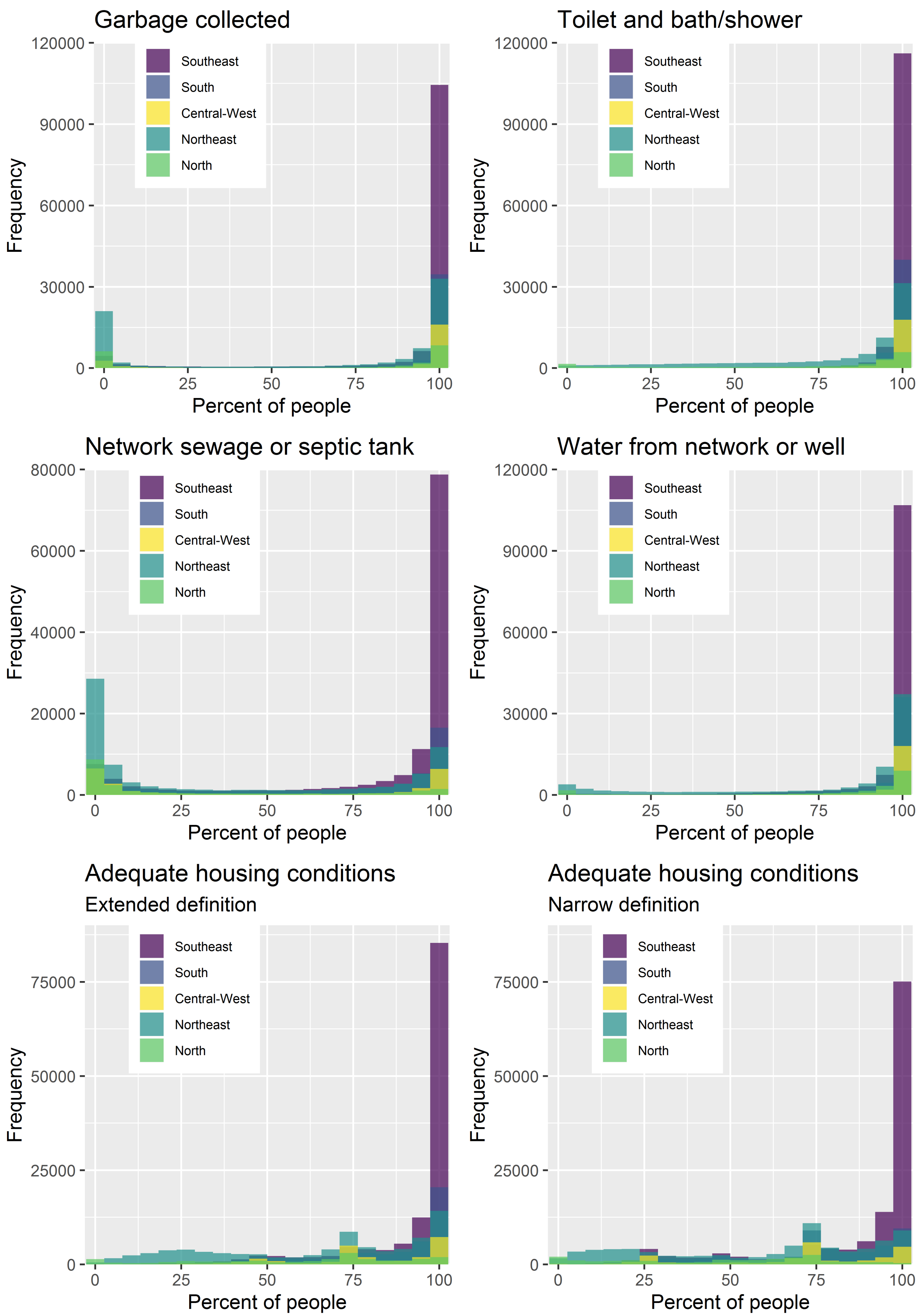


Figure 6: Histograms of individual housing conditions, by region

differences are quite distinct.

The individual housing condition variables have a weak to moderate correlation with each other, with the highest correlations between access to network water and garbage collection (r = 0.74) and the lowest between access to electricity and sewage (r between 0.23-0.28, see Table A.2). For these reasons we have decided to combine the four variables on access to toilet and shower/bath, garbage collection, water and sewage into a single housing conditions variable. We have combined them by simply taking the mean of the four variables. This would give the average percentage of people in a census sector experiencing any of the conditions. We have considered two alternative definitions, the narrow definition of housing characteristics includes the percentage of people with access to network sewage and water supply only. The extended definition includes people with access to the network or a septic tank for sewage disposal and to the network or a well for water supply.

The distributions and summary statistics for both of these variables are given in Table 8 and Figure 6. The narrower definition has a lower mean and better variation by definition, but the extended definition has a better conceptual fit to deprivation, particularly in rural areas. The histograms for both variables show that in the South and Southeast, nearly all census sectors have a very high percentage of people with adequate housing. In other regions there is much more variation in housing conditions. The summary statistics for the extended definition by region also show clear regional variation. In the Southeast the mean for adequate housing is 91.5%, in the South 84.9% and in the Central West 79.2%, but in the Northeast it is 64.5% and in the North 59.8%

5.4 Neighbourhood characteristics

The census also collected information on neighbourhood characteristics, such as on street lighting and road surfacing. These data are provided in five tables, *Entorn01* through *Entorno05*. Unfortunately, these data are predominantly collected for urban areas. Table 9 gives the percentage of census sectors for which data are collected by the urban-rural classification and by the five regions. While almost all urban areas have environmental data, only 2% of rural areas have this information present. As Northern regions have more rural areas, the data coverage is lower in the North and Northeast (58% and 61%) and higher in the South and Southeast (72% and 83%), where there are fewer rural areas. Because the data only cover urban areas, we have not used environment as a domain in the deprivation measure. However, for the 73% of sectors for which we have data, we compare the environment variables to the other deprivation domains to gauge the impact it may (or may not) have on the overall deprivation measure.

Information is collected for 10 aspects of the local environment, including signs for street names, street lighting and paving, the presence of a sidewalk for pedestrians, a raised curb, storm drains, wheelchair ramps or sloping sidewalk curbs, green space, open sewage drainage and litter on the streets. The data provide both the number of permanent private households and the number of people in permanent private households in a census sector experiencing any of the listed neighbourhood characteristics. Of all the variables provided, we concluded that street lighting, road surfacing (such as tarmac), presence of sidewalks and storm drains

Table 9: Coverage of environment data by urban-rural classification and region

Sector location	Data collected	Not collected
Rural	2.1	97.9
Urban	94.3	5.7
Region		
North	60.8	39.2
Northeast	58.2	41.8
Central-West	77.1	22.9
South	71.6	28.4
Southeast	83.2	16.8
Total	72.8	27.2

were the best variables to capture deprivation. We also considered using the presence of open sewage and litter on the streets, but decided that these were too similar to the housing condition variables on garbage collection and access to sewage disposal that we already planned to include in the measure.

For all variables we used the number of people in permanent private households experiencing any condition as the numerator. As for the housing conditions, the number of people in all private households is used as the denominator. This means that all people in improvised households were counted as not having access to street lighting, road surfacing, sidewalks or storm drains in their neighbourhood. People in collective residence were excluded from the measure.

Table 10 shows the summary statistics and the top four panels in Figure 7 the histograms for the four neighbourhood variables. The distributions are very skewed and in two cases also bimodal. Most sectors, regardless of the region, have a very low percentage of people with no street lighting, road surfacing or sidewalks in the neighbourhood. Compared to other variables, storm drains are more likely to be absent from the neighbourhood, and this is much more likely to be the case in the Northern regions compared to the Southern regions.

The correlations between the four variables are quite weak, with the correlation coefficient at 0.22 between street lighting and storm drains and 0.42 for paving and storm drains (see Table in Appendix A). As may be expected, the correlation is strongest between presence of street pavement and sidewalks ($r = 0.72$).

As with housing conditions, we do not have a count of people who lack access to a combination of these neighbourhood characteristics and to combine these we have decided to use the

Table 10: Summary statistics for housing condition variables

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD	Missing
No Streetlight	0	0.0	0.0	5.1	3.2	100	13.9	89521
No pavement	0	0.0	1.3	19.8	28.9	100	31.0	89521
No sidewalk	0	0.0	11.6	32.8	68.2	100	38.1	89521
No storm drain	0	21.9	67.8	59.0	98.3	100	37.8	89521
Average environment	0	10.69	25.00	29.15	44.06	100	23.4	89521
Average environment, excluding lighting	0	13.78	32.86	37.17	56.58	100	29.0	89521

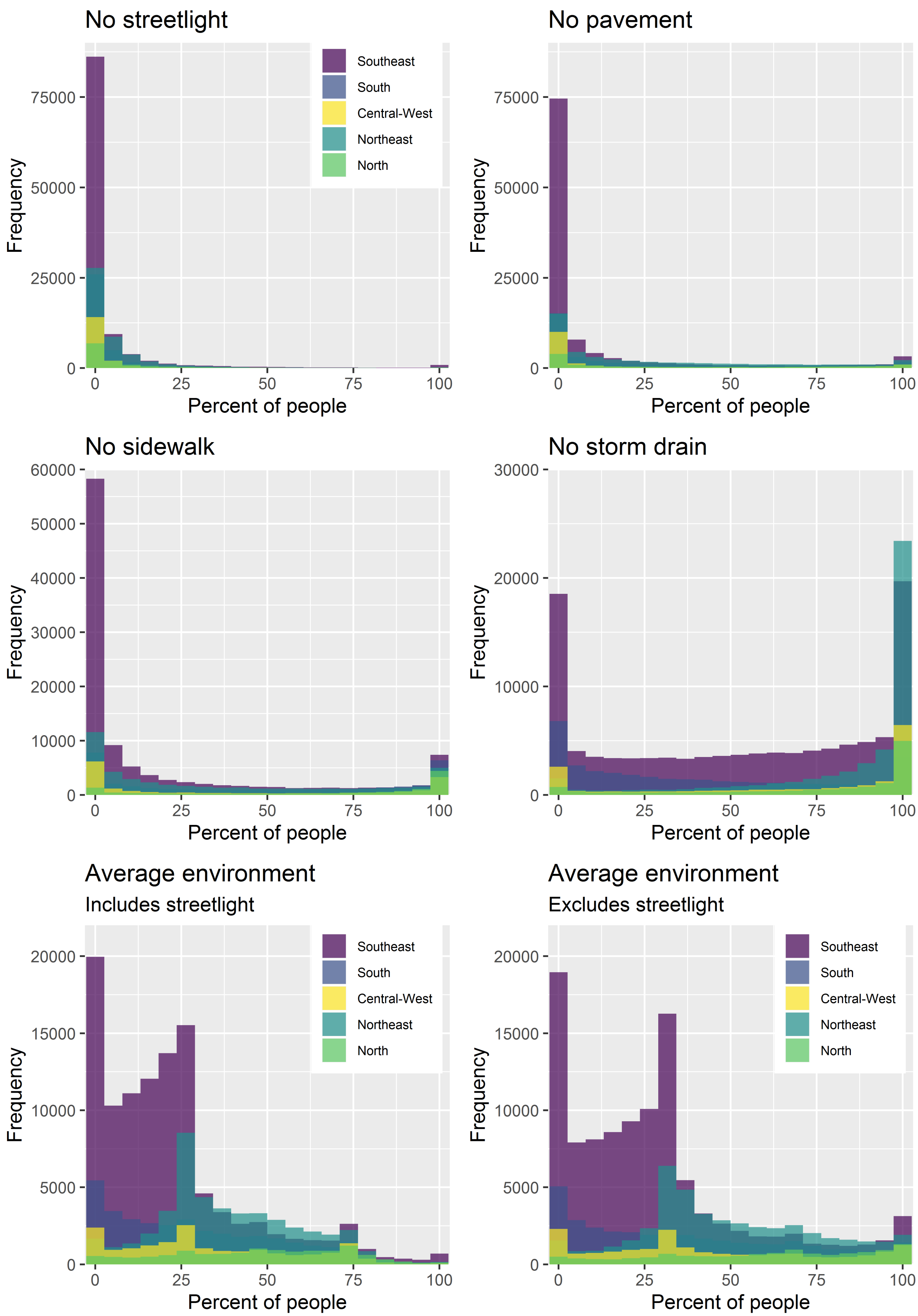


Figure 7: Histograms of environment characteristics, by region

arithmetic mean again. Table 10 provides summary statistics for the average environment variable and Table A.3 in Appendix A the correlations between the individual variables. Street lighting has the weakest correlation with the average environment indicator and also has quite low variation in general. For these reasons, a case could be made for excluding it from the average environment indicator. Excluding street lighting has little impact on the average neighbourhood conditions and the correlation coefficient between the two averages is 0.99. Distributions for the two environment variables, including and excluding streetlight, are shown in the bottom two panels in Figure 7.

5.5 Comparing indicators and domains

We have considered 14 different indicators available from the census, reflecting material wealth, education, housing conditions and neighbourhood characteristics. For income, the choice of variables available to develop a deprivation measure was quite varied, but for literacy rates and housing conditions it was much more limited. The different literacy and housing variables also had low variability in values across Brazil compared to data on income. Neighbourhood characteristics were predominantly only recorded for urban areas and for this reason had to be excluded from the measure.

Despite some limitations, the three domains together meet many of the desirable qualities of deprivation indicators: they are provided consistently for small-areas across Brazil, are conceptually sound and capture different aspects of deprivation. Having selected a small number of potential deprivation indicators for each of the three domains, we will now compare each of these to the other domains. The final choice will be made by taking account of correlations between the indicators across Brazil and by regions. While income, literacy and housing are different aspects of material deprivation, they should all still be fairly well correlated if they measure the same phenomenon.

Table 11 shows the correlations between the seven indicators. It is evident that literacy and income have a much stronger correlation to each other compared to the housing variables. Of the two housing conditions variables, the extended definition correlates better to both income and literacy compared to the narrower definition.

The correlations vary between regions and are generally stronger in the Northern regions, compared to the Southern regions. Correlation coefficients between income and literacy vary between -0.73 and -0.76 in the North, but are between -0.56 and -0.71 in the South. Notably, household income below 1/4 MW has a weaker correlation to literacy in the South. The extended housing conditions variable correlates better with literacy and income in all regions compared to the narrow definition and this difference is more pronounced in the South.

We found similar patterns in the correlations in urban and rural areas. The narrower housing conditions definition had a weaker correlation to the other domains in both urban and rural areas. Household income below 1/2 MW was better correlated to literacy and housing in urban areas compared to the lower cut-off of 1/4 MW. Literacy variables all had similar correlation coefficients with other variables in both urban and rural areas.

Based on these correlations we prefer to use the extended housing conditions variable over

Table 11: Correlations between indicators, across Brazil and by region

Variable	Income < 1/4	Income < 1/2	Literacy 5+	Literacy 7+	Literacy 7-9 & 50+	Housing, extended	Housing, narrow
Income < 1/4	1.00	0.92	-0.81	-0.80	-0.79	-0.73	-0.65
Income < 1/2	0.92	1.00	-0.82	-0.80	-0.84	-0.72	-0.66
Literacy 5+	-0.81	-0.82	1.00	1.00	0.96	0.75	0.67
Literacy 7+	-0.80	-0.80	1.00	1.00	0.95	0.74	0.67
Literacy 7-9 & 50+	-0.79	-0.84	0.96	0.95	1.00	0.71	0.64
Housing, extended	-0.73	-0.72	0.75	0.74	0.71	1.00	0.94
Housing, narrow	-0.65	-0.66	0.67	0.67	0.64	0.94	1.00
	Northern regions combined						
Income < 1/4	1.00	0.91	-0.75	-0.74	-0.73	-0.74	-0.71
Income < 1/2	0.91	1.00	-0.76	-0.74	-0.79	-0.70	-0.67
Literacy 5+	-0.75	-0.76	1.00	1.00	0.95	0.71	0.67
Literacy 7+	-0.74	-0.74	1.00	1.00	0.94	0.70	0.65
Literacy 7-9 & 50+	-0.73	-0.79	0.95	0.94	1.00	0.67	0.63
Housing, extended	-0.74	-0.70	0.71	0.70	0.67	1.00	0.96
Housing, narrow	-0.71	-0.67	0.67	0.65	0.63	0.96	1.00
	Southern regions combined						
Income < 1/4	1.00	0.88	-0.58	-0.58	-0.56	-0.47	-0.40
Income < 1/2	0.88	1.00	-0.71	-0.70	-0.70	-0.53	-0.48
Literacy 5+	-0.58	-0.71	1.00	0.98	0.94	0.57	0.51
Literacy 7+	-0.58	-0.70	0.98	1.00	0.94	0.59	0.53
Literacy 7-9 & 50+	-0.56	-0.70	0.94	0.94	1.00	0.50	0.43
Housing, extended	-0.47	-0.53	0.57	0.59	0.50	1.00	0.91
Housing, narrow	-0.40	-0.48	0.51	0.53	0.43	0.91	1.00

the narrower definition as it has a stronger relationship with other deprivation domains. Particularly important is the fact that in Southern regions the correlation coefficients for the narrower definition falls below 0.5, reflecting a relatively weak relationship with other deprivation indicators.

For income, the cut-off point below 1/4 MW works better for Northern regions and the cut-off 1/2 MW better for Southern regions, reflecting the well known economic differences between the North and South in Brazil. Because the lower 1/4 MW cut-off is only slightly better correlated with literacy and housing conditions in the North, but has a much worse correlation to the other domains in the South, we have decided to use household income below 1/2 MW in the deprivation measure. Using the higher cut-off increases variability in the income domain in the Southern regions and is thus better able to distinguish between small-areas.

As for literacy rates, using either the age group 5 and above or 7 and above makes almost no difference to the results. We have chosen to use age 7 as the cut-off as this reflects the mandatory school age in Brazil and is therefore a more reasonable indicator of access to education and therefore also of deprivation. Literacy for ages 7-9 and 50+ has a similar relationship to income compared to the other literacy measures, but a slightly weaker relationship to housing conditions. The variable does, however have better variability

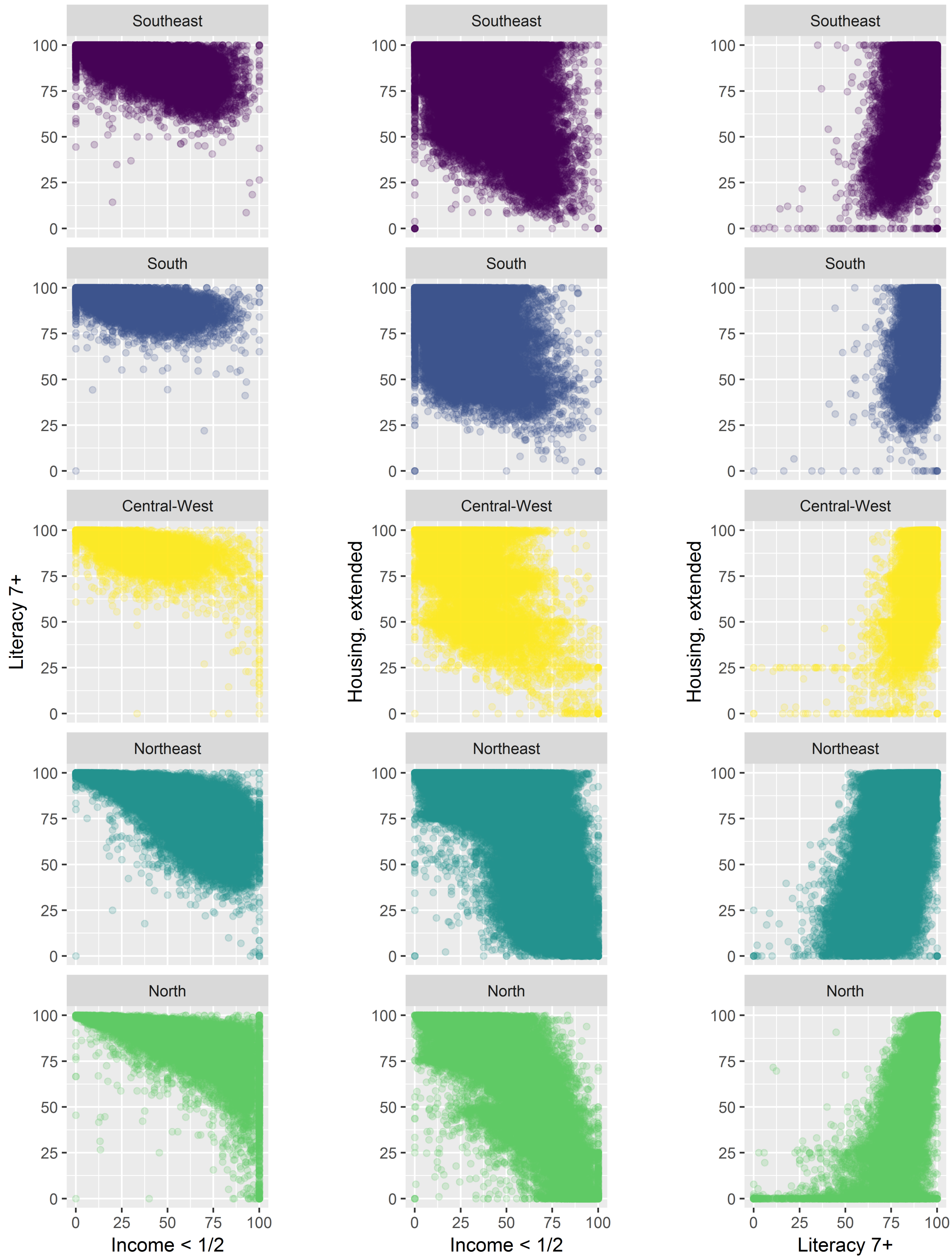


Figure 8: Correlation of deprivation indicators by region

compared to the other two literacy rates. In the end, we decided to use the simpler age group 7+ as the variable that combines ages 7-9 with 50 and over is based on a smaller population.

The scatter plots in Figure 8 show the relationships between the three selected variables for the income, education and housing domains. The plots show a substantial amount of scatter and a number of areas have very different deprivation levels for the different domains. Thus, while the correlations are reasonably strong, there is a lot of variability in deprivation across the domains.

Comparing the indicators from one domain to other domains across Brazil and within regions has helped us narrow our choice down to three deprivation indicators to be included in the measure. Our decision was based on the desirable criteria that the indicators from different domains should be well correlated and this relationship should be consistent across different regions and for both rural and urban areas.

We also tried to choose indicators with higher variability and in the case of household income, the cut-off point 1/2 MW achieves this better than income below 1/4 MW. For housing and literacy rates we ended up choosing a variable with somewhat lower variability, but better overall coverage across Brazil for literacy, and better conceptual fit and correlation to other domains for housing. To conclude, the three variables included in the final measure are:

- percent of households with *per capita* income $\leq 1/2$ minimum wage;
- percent of people not literate, aged 7+;
- average of the percent of people with inadequate access to sewage, water, garbage collection and no toilet and bath/shower.

5.5.1 Comparison to neighbourhood characteristics

The average neighbourhood characteristics variables have a moderate correlation to the income and literacy domains, but the correlation between the neighbourhood and housing domains is stronger. Of the individual neighbourhood characteristics, no street lighting has the weakest correlation to any other deprivation indicator and removing this from the average neighbourhood characteristics variable has not impact on the correlations with the other deprivation variables. Of the individual variables, road surfacing and sidewalks have the strongest correlations to income, literacy and housing conditions.

Table 12: Correlations of environment variables to other domains

	Income < 1/2	Literacy 7+	Housing, extended
Average environment	0.56	-0.47	-0.64
Average environment, excluding lighting	0.56	-0.47	-0.63
No streetlight	0.24	-0.19	-0.36
No road surfacing	0.45	-0.37	-0.57
No sidewalk	0.46	-0.36	-0.60
No storm drain	0.46	-0.42	-0.37

6 Combining domains

There are many options available when it comes to combining the indicators into a measure (Morris and Carstairs 1991), the most common of these are outlined below. While there is no single commonly agreed correct method, previous research has outlined some desirable criteria for combining indicators. These include ensuring that the indicators are measured on the same scale and share similar distributions (Noble et al. 2006). This should result in a measure where no single indicator implicitly drives the result, such as by having a very large impact on the outcome by virtue of how it is measured.

Sum of standardized scores. One of the earliest and easiest methods to combine variables is to calculate the standardized scores (z-scores) for each variable and then simply adds these to get the measure (Carstairs and Morris 1991; Townsend, Phillimore, and Beattie 1988). Other standardization methods include using the minimum and maximum values of variables (UNDP 2018; IPEA 2019; Vasquez 2016). Calculating z-scores is a fairly straightforward method and still quite commonly used (Vasquez 2016; Norman, Berrie, and Exeter 2019). In most cases equal weights have been used for the different indicators (Green et al. 2018; Carstairs and Morris 1991; Townsend, Phillimore, and Beattie 1988), but this is not a requirement and different weights can be used (Jarman 1983). Standardization puts variables with different means on the same scale, meaning indicators with very different means will still have the same impact on the overall measure. However, z-scores do not change the distribution of the variables and for this reason outliers or heavily skewed variables may have a disproportionate effect on the measure.

Principal component or factor analysis. Principal component analysis (PCA) and factor analysis are very popular methods of combining variables into a measure (Lalloué et al. 2013; Messer et al. 2006; Bonfim et al. 2009). The benefits of these and other similar statistical methods include a relatively straight forward implementation and that they do not require the researcher to make any judgement about the weight of each variable (Sánchez-Cantalejo, Ocana-Riola, and Fernández-Ajuria 2008). When PCA is based on the correlation matrix it is also not dependent on scale.

Criticisms of these method include the data driven assignment of weights to variables/domains. This may mean that variables that are more strongly correlated with each other have more impact (higher weight) on the measure and other variables will have a weaker impact (lower weight) on the measure (Bonfim et al. 2009). While this may be desirable, it is important to be aware of this and to discuss how this impacts the measure.

Transformation to ranks and exponential distribution. The many UK deprivation measures such as the SIMD (The Scottish Government 2012) or the English IMD (Smith et al. 2015) rank deprivation variables and then transform these ranks into exponentially distributed domains scores. The domain scores are then weighted and combined. The benefits of this include that many very different variables (with different distributions and summary statistics) are standardized and transformed into a common distribution before they are combined. The exponential distributions (and the long right-hand tail of the distribution) also allow for the better identification of the most deprived areas.

This method has been criticized for imposing an artificial distribution on variables and simplifications to the methodology could be made with little impact on the resulting deprivation indices (McConnachie and Weir 2005). In the case of Brazil, many of the variables used are already exponentially distributed, share the same maximum and minimum values and have similar means and medians. For this reason we have not seen the need to use ranks to transform the variables into exponentially distributed scores.

Geometric mean. The geometric mean approach is used by the United Nations when creating the Human Development Indices (UNDP 2018) and also the Municipal Human Development Index (MHDI) used in Brazil (The Atlas 2019a). It is a simple method and, unlike the arithmetic mean, takes account of the scales of the different variables. However, using the geometric mean is problematic when the variables take on zero values. This will lead to the geometric mean itself being zero, thereby cancelling out all other aspects of deprivation. In our case, there are a lot of census sectors where one of the variables take on the value zero. For housing conditions about 15% of all census sectors are coded zero. The lack of deprivation on a single domain should not, however, indicate complete lack of deprivation across all domains and for this reason we have opted not to use the geometric mean.

Arithmetic mean. Using the arithmetic mean, i.e. averaging across untransformed variables, is often not used for combining variables. It is easily impacted by extreme values and when variables have very different means or are measured on different scales, it leads to different domains having a very different impact on the measure. However, in some cases averaging may be appropriate. The municipal Social Vulnerability Index (IVS) for Brazil uses the arithmetic mean on domain scores that are standardized using the minimum and maximum values (IPEA 2019). In our case, many of the variables available have similar means, distributions and are all measured on the same scale. For this reason, a simple average of the three domains may be appropriate. There are also benefits to using the arithmetic mean, such as easy interpretation of the results. The resultant score from can be interpreted as the mean level deprivation, but for other methods the numeric values themselves have no meaning.

One of the main differences between using the mean (either arithmetic or geometric) versus using standardized scores or PCA is that a score based on either of the means is calculated independently for each area without using any information from other areas. If the variables take on values 3, 4, and 5, the means for the area are 4 for the arithmetic and 3.9 for the geometric mean. The z-scores, however, use the standard deviation of the variable to adjust for the spread in the data. Values that are different from the mean will have a lesser impact on the overall score if the standard deviation is high compared to if the standard deviation is low. If the different deprivation indicators have different levels of variation, using standardization will produce a different relative ranking of areas compared to the arithmetic or geometric mean. PCA, calculated using the correlation matrix, will standardize the variables like the z-score method.

There is no single correct method of combining variables into a single measure. The decision depends on value judgements and desirable criteria set out by the researchers. Empirically, there is little evidence to suggest that the method of combining variables has a substantial impact on the measure (Fukuda, Nakamura, and Takano 2007; Krieger et al. 2002; Lòpez-De Fede et al. 2016). We decided that a deprivation score for a single area should reflect variation in the data. This is because we are measuring relative and not absolute deprivation. Standardization, such as using z-scores or PCA, will score any single area relative to the overall variation in data. We have opted for the z-score method as it allows an explicit assignment of weights and is more easily replicated in the exact same manner for different geographic area levels and time-points. However, we show below that the difference between the z-score method and PCA is minimal. We also compared the measure based on the z-scores to a simple arithmetic mean of the indicators and found little difference in the results.

6.1 BrazDep 2010

The BrazDep 2010 deprivation measure for Brazil was calculated by combining z-scores of the three deprivation indicators identified earlier: percent of households with income below 1/2 MW; percent of people aged 7 and above who are not literate; percent of people with inadequate access to sewage, water and garbage collection, and without bathroom/shower and toilet. The z-scores z for a variable x is calculated using the formula $z = (x - \mu)/sd$, where the mean μ and standard deviation sd for the individual indicators are population weighted (Brown et al. 2014). The calculations were done using the `zscore` function in R package `SocEpi` (Allik 2019). The z-scores were weighted equally by simply summing these into a single deprivation measure.

Table 13 shows the summary statistics and Figure 9 the distribution of the BrazDep measure. The high positive values of the BrazDep 2010 measure indicate high and the low negative values low deprivation. Values close to zero indicate average deprivation. Beyond that, the numeric values themselves have no substantive meaning. The deprivation measure is exponentially distributed, with a large proportion of areas having a low deprivation score and a smaller number of areas experiencing very high deprivation.

A comparison of the summary statistics across regions shows that the Northern regions have higher mean and median deprivation compared to the Southern regions. The Central-West regions falls between the North and South in terms of average deprivation. All regions have very deprived and not at all deprived areas, as shown by the minimum and maximum values, but the differences between North and South are very distinct and clearly evident from the histogram. Differences in deprivation between urban and rural areas are similar to regional

Table 13: Summary statistics of the BrazDep measure, across Brazil, by region and by urban rural classification

	Summary statistics							N of sectors		% sectors
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD	Available	Missing	Available
Brazil	-3.12	-2.08	-0.76	0.21	1.67	15.59	2.96	303218	6902	97.77
Northeast	-3.12	0.25	2.76	2.88	5.60	15.59	3.21	79061	1287	98.40
North	-3.12	-0.25	1.71	2.33	4.52	15.59	3.36	20912	453	97.88
Central-West	-3.12	-1.85	-0.61	-0.31	0.83	15.59	2.01	23418	501	97.91
Southeast	-3.12	-2.40	-1.66	-1.17	-0.49	14.39	1.71	130003	2806	97.89
South	-3.12	-2.42	-1.56	-1.08	-0.10	11.81	1.67	49824	1855	96.41
Urban	-3.12	-2.32	-1.43	-0.91	-0.00	15.15	1.86	232353	4361	98.16
Rural	-3.12	1.40	3.94	3.87	6.22	15.59	2.96	70865	2541	96.54

differences. The means and medians for rural areas are much higher, indicating higher deprivation in rural areas compared to the urban areas. The histogram by the urban-rural classification in Figure 9 shows the majority of urban areas have low deprivation and the tail end of the distribution is dominated by the rural areas.

The table also includes the number and percent of census sectors for which the measure could be calculated and the number of sectors where there was not sufficient data available for calculating the measure. The measure is provided for 97.8% of sectors across Brazil and the variation of this percentage across regions and urban-rural classification is quite small.

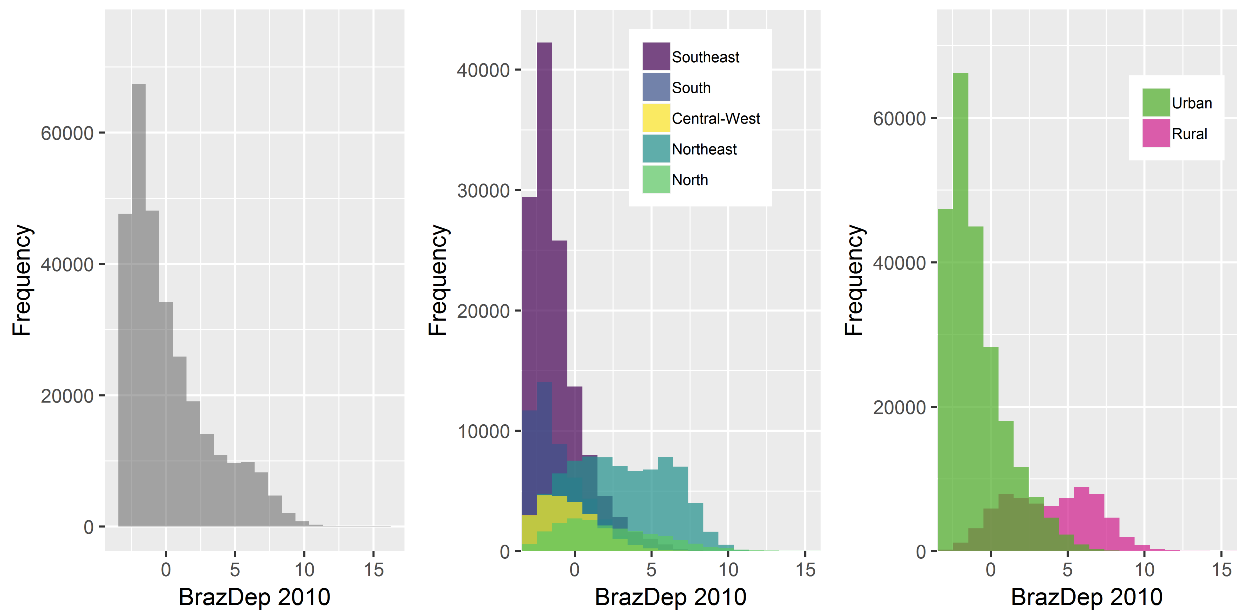


Figure 9: Histograms of BrazDep 2010 measure, across Brazil and by regions

6.2 Deprivation categories

It is often useful to group areas into deprivation categories, such as quintiles (5ths), deciles (10ths) or vigintiles (20ths). We have calculated population weighted deciles and quintiles for the BrazDep measure. Population weighted quintiles group small areas into five categories, from least deprived to most deprived, such that each will include 20% of the population. Deciles group small areas into 10 groups, each including 10% of the population.

Table 14 shows the distribution of census sectors and population by deprivation deciles for the five regions. The first decile indicates the least and the 10th decile the most deprived areas. Since across Brazil each decile includes 10% of the population, any variation from this within the regions will indicate higher or lower deprivation compared to the whole of Brazil. In the two Northern regions the proportion of both people and census sectors in the most deprived areas (e.g. 9th and 10th deciles) is much higher compared to the Southern regions, meaning that deprivation is higher in the North and Northeast. South and Southeast include the highest percentage of people in the least deprived 1st and 2nd quintile and are less

Table 14: Distribution of census sectors and population by decile, five regions

Region	Low deprivation 1	2	3	4	5	6	7	8	9	High deprivation 10	Total
	Percent of census sectors										
Northeast	2.6	2.1	2.3	3.0	4.3	6.1	8.9	13.3	20.4	36.7	100
North	1.8	2.5	3.1	4.0	5.8	8.4	12.3	15.5	20.0	26.8	100
Central-West	9.4	7.2	8.1	9.2	10.8	12.6	15.8	15.2	9.3	2.4	100
Southeast	15.8	14.4	13.5	13.0	11.7	10.3	8.4	6.7	4.5	1.6	100
South	16.4	14.1	11.9	10.7	10.4	10.1	10.3	10.1	5.4	0.5	100
	Percent of people										
Northeast	2.6	2.3	2.7	3.5	5.1	7.3	10.6	16.0	22.4	27.6	100
North	1.7	2.5	3.3	4.3	6.3	9.8	14.4	17.9	19.6	20.2	100
Central-West	8.9	8.0	9.2	10.4	12.4	14.6	17.2	12.2	5.6	1.5	100
Southeast	14.4	14.9	15.1	14.7	13.1	10.9	7.8	5.1	3.0	1.1	100
South	16.9	16.1	13.8	12.1	11.5	10.4	8.9	6.9	3.2	0.3	100

deprived. The Centre West region has the highest percentage of people in deciles 5 through 8, indicating average to moderate deprivation compared to the other regions of Brazil.

The distribution of census sectors and people by deciles in Table 15 show the differences in deprivation levels for urban and rural areas very clearly. 47% of all rural sectors and 53% of people in rural areas live in the most deprived (10th) decile. In urban areas, both numbers are only at 2 percent. Less than 1% of sectors and people in rural areas are in the least deprived 1st and 2nd deciles. In urban areas these percentages are 27% and 24% for sectors and people respectively in the least deprived first and second decile combined.

The distribution of sectors and population by deciles for the different sector types in Table 16 also shows clear relationships between deprivation and sector type. Some sector types, such as rural settlement projects or indigenous villages, have a very high proportion of deprived sectors and people in these sectors. 84% of people in indigenous villages and 65% of people in rural settlement projects live in the most deprived (10th) decile. Special agglomerates also have a relatively large proportion of people in the moderately deprived 5th to 8th deciles and a very low proportion of people in the least deprived areas. Sectors that are classified as military bases have an above average proportion of least deprived sectors and 38% of people in these areas are in the least deprived decile.

Table 15: Percent of census sectors and people by decile, urban and rural areas

Sector type	Low deprivation 1	2	3	4	5	6	7	8	9	High deprivation 10	Total
	Percent of census sectors										
Rural	0.2	0.3	0.6	1.1	2.0	4.2	8.6	15.0	21.1	47.0	100
Urban	14.3	12.6	11.8	11.6	11.2	10.8	10.1	8.8	7.0	1.8	100
	Percent of people										
Rural	0.1	0.3	0.5	1.0	1.8	3.6	6.9	12.1	20.5	53.2	100
Urban	11.8	11.8	11.8	11.7	11.5	11.2	10.6	9.6	8.1	2.0	100

Table 16: Percent of census sectors and people by decile, census sector types

Sector type	Low deprivation 1	2	3	4	5	6	7	High deprivation 8	9	10	Total
	Percent of census sectors										
Common or non-special sector	11.7	10.4	9.7	9.3	8.9	8.6	9.1	9.9	10.2	12.3	100
Special agglomerate	0.2	0.7	2.4	7.1	14.8	22.4	23.0	19.0	8.8	1.5	100
Barracks, military bases, etc.	38.6	15.9	9.1	6.8	9.1	6.8	9.1	2.3	0.0	2.3	100
Special accommodation, camps, etc.	33.3	8.3	8.3	8.3	0.0	8.3	0.0	0.0	25.0	8.3	100
Indigenous village	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.7	8.0	89.9	100
Prisons, etc.	9.5	9.5	9.5	19.0	14.3	4.8	4.8	9.5	9.5	9.5	100
Orphanages, convents, hospitals, etc.	1.7	8.5	3.4	3.4	8.5	18.6	16.9	16.9	15.3	6.8	100
Rural settlements projects	0.0	0.0	0.0	0.1	0.1	0.4	0.8	4.7	27.4	66.6	100
	Percent of people										
Common or non-special sector	10.7	10.7	10.6	10.3	9.8	9.3	9.2	9.4	10.0	10.3	100
Special agglomerate	0.1	0.6	2.0	6.4	14.2	22.3	23.6	20.0	9.4	1.4	100
Barracks, military bases, etc.	38.3	20.4	7.4	3.0	12.8	7.6	3.6	2.1	0.0	4.7	100
Special accommodation, camps, etc.	66.3	6.7	0.4	1.6	0.0	2.6	0.0	0.0	19.7	2.8	100
Indigenous village	0.0	0.0	0.0	0.0	0.0	0.0	1.1	2.5	12.8	83.6	100
Prisons, etc.	22.4	6.8	10.5	19.8	17.8	0.6	2.0	12.0	4.0	4.0	100
Orphanages, convents, hospitals, etc.	0.5	6.9	7.3	1.1	8.5	17.5	9.5	15.0	21.1	12.6	100
Rural settlements projects	0.0	0.0	0.0	0.1	0.1	0.8	1.2	3.9	29.0	64.8	100

6.3 Comparison to alternative methods

To understand what effect the methods of combining variables have on the relative deprivation of census sectors we have compared the BrazDep measure to scores calculated using PCA and the arithmetic mean. For the former method we used the `principal` function in the R package `psych` (Revelle 2018). The PCA is based on the correlation matrix as opposed to the variance-covariance matrix. Using the correlation matrix ensures that the variables are standardized before the principal component is calculated and variables with higher variance (e.g. income and housing in our case) will not dominate the principal component. We derived a single component based on the three deprivation indicators identified earlier and this is compared as an alternative to the BrazDep deprivation measure.

The results of the PCA analysis are shown in Table 17. The loadings of the three variables on the principal component are all very similar. The component itself explains a large proportion of variance in the three indicators and tests confirm that a single component is sufficient to explain variation in the data adequately.

The correlations between the three alternative measures across Brazil are very strong (r $\geq$ 0.99). The scatter plots in Figure 10 visualize these relationships. There is almost no difference between the BrazDep measure calculated from the z-scores and the PCA method. More differences in relative deprivation are visible when the arithmetic mean is compared

Table 17: Principal component analysis

Variable	Loading
Household income $\leq$ 1/2 MW	0.920
Percent not literate aged 7+	0.930
Housing conditions, extended	0.894
Proportion explained	0.837

to the PCA and z-score based scores, but even these are quite small. Across regions the correlations between the alternative methods are just as high.

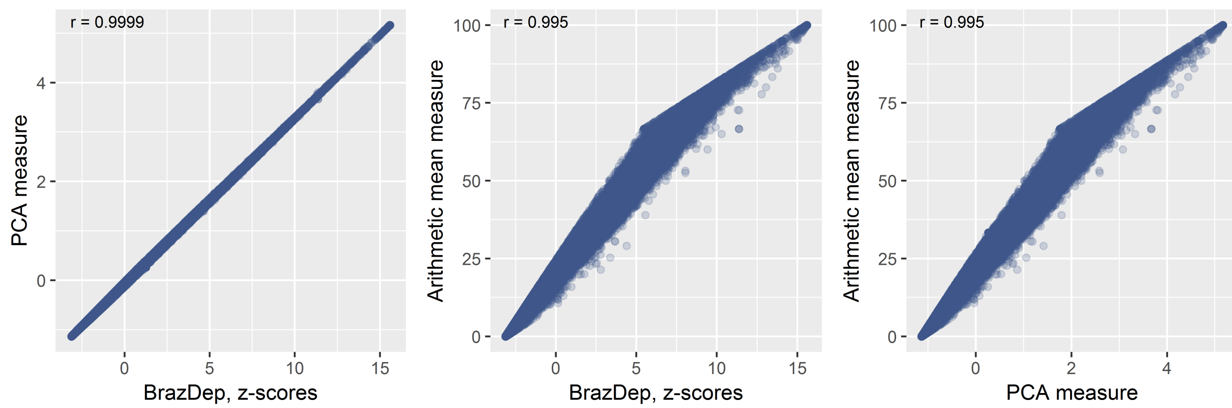


Figure 10: Relationships between alternative measures

A comparison of deciles gives the same results. Table 18 presents a cross tabulation between the BrazDep deciles and deciles for a measure that uses the arithmetic means. Of all census sectors, 83.6% are on the diagonal, meaning that they have the same decile value and 16.4% are one decile above or below the diagonal. Only 130 sectors (0.04%) are more than one decile apart. The differences in deciles between the BrazDep and PCA method are even smaller and all sectors are either on the diagonal or only one decile apart. In conclusion, the three different methods of combining variables give very similar results, especially for z-scores and PCA, which both standardize the variables.

Table 18: Comparison of deciles using different combination methods

		BrazDep, z-scores									
		Low deprivation									High deprivation
		1	2	3	4	5	6	7	8	9	10
Arithmetic mean	1	31382	2000	1	0	0	0	0	0	0	0
	2	1885	25018	2647	5	0	0	0	0	0	0
	3	4	2472	22267	3158	1	0	0	0	0	0
	4	0	16	2995	21315	3334	1	0	0	0	0
	5	0	0	25	3105	21256	3118	1	0	0	0
	6	0	0	0	29	3062	21813	3034	0	0	0
	7	0	0	0	1	28	3024	23219	3174	1	0
	8	0	0	0	0	0	7	2965	24996	3174	0
	9	0	0	0	0	0	0	9	2567	26445	2105
	10	0	0	0	0	0	0	0	1	1765	35793

7 Validating the deprivation measure

Validation of deprivation measures is quite difficult, as there is often no gold standard to test the measures against. Subsequently, examples of and guidance for validation are rare (Pampalon et al. 2014; Carr-Hill and Chalmers-Dixon 2005). One option for validating the measure is to compare it to other similar measures developed at the census sector level for specific states or regions. We have access to two such measures, one for the municipality of Belo Horizonte and one for the state of São Paulo. If the BrazDep measure correlates well to these measures in the two locations of Brazil, then it may also be valid for other areas of Brazil.

To provide validation for the whole of Brazil, we have also compared the BrazDep measure to similar measures developed for municipalities. We have compared our measure to the IVS and to the MHDI, both provided for the whole of Brazil. If the measures are correlated, we should have more validity that the developed measure is valid also for other areas of Brazil.

One of the best method for validation may be to test how well the deprivation measure can explain outcomes of interest, such as mortality or ill health. Comparing the predictive power of different indicators or measures can highlight those that are less able to explain variation in health outcomes (Allik et al. 2016; Krieger et al. 2003). Unfortunately, mortality and health data in Brazil are currently not available at the census sector level. However, we are able to compare our measure to life expectancy, mortality and survival at the municipality level, lending more validity to the measure.

7.1 Comparison to other census sector level measures

7.1.1 Belo Horizonte

The municipality of Belo Horizonte has developed a census sector level Health Vulnerability Index (*Índice de Vulnerabilidade à Saude*, IVS) (Prefeitura Belo Horizonte 2013). The first version of this was released in 1998 and the 2012 IVS is the third version. The 2012 IVS is calculated using the 2010 census, combining 18 indicators related to housing conditions (sanitation), household crowding, literacy, income, ethnicity and neighbourhood. The indicators are standardized using minimum and maximum values and then weighted. The indicators have different weights, with housing condition having the highest and crowding the lowest weights. Weights are based on expert opinion. The indicators are combined into a single index (the IVS), but also into two domains (sanitation and socioeconomic domains).

The combined index is used to develop the four health vulnerability categories:

1. Medium risk – census tracts within 1/2 standard deviation of the mean value of the index (mean +/- SD 0.5).
2. Low risk – tracts with IVS values lower than medium IVS group.
3. High risk – sectors with values above medium IVS and up to 1.5 standard deviation above the mean value of the index.
4. Very high risk – sectors with values above the high IVS group.

Table 19: Distribution of census sectors and population by IVS and BrazDep Quintiles, Belo Horizonte

BrazDep	Low				High
	1	2	3	4	5
Sectors	52.5	26.7	16.0	4.6	0.2
Population	52.6	27.4	15.8	4.1	0.1

IVS	Low	Medium	High	Very high
Sectors	34.7	38.1	19.2	7.9
Population	33.7	39.9	19.1	7.3

The index is provided for 3830 of the 3905 census sectors in Belo Horizonte. The BrazDep measure is provided for the same sectors and also for an additional 11 areas. Figure 11 gives the histograms of the IVS and BrazDep measures and Table 19 gives the distribution of CS and population by the categorical IVS measure and BrazDep quintiles. Note that the BrazDep quintiles were calculated using data for the whole of Brazil and for this reason the quintiles do not all have exactly 20% of the population in them. Over half of all sectors and people in Belo Horizonte are in the least deprived 1st quintile and a further 27% are in the second quintile. Less than 1% of sectors and people are in the most deprived quintile. This skewed distribution of sectors and people by the BrazDep quintiles reflects the relatively affluent nature of Belo Horizonte in comparison to other municipalities in Brazil. The IVS also provides similarly skewed distributions for census sectors and the population. A third of all sectors and people are in the low and about another third in the medium vulnerability category. Only 8% of sectors and people are in the very high health vulnerability group.

As the two measures use similar indicators, we would expect them to be highly correlated. The relationship between the IVS and the BrazDep can be seen in the scatter plot in Figure 11. While there are some outliers, the relationship is quite clear and strong, with the correlation coefficient equal to 0.93.

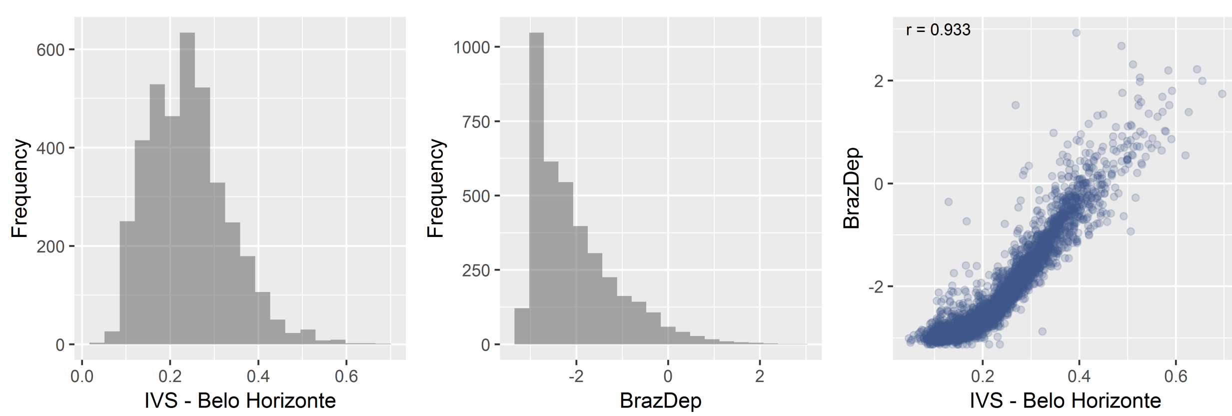


Figure 11: Belo Horizonte census sector measures

The comparison of the categorical measures also shows a strong relationship. Table 20 gives the BrazDep quintiles by health vulnerability categories (column percentages). Nearly all (98%) of the low vulnerability sectors are in the least deprived first quintile and a similar percentage of areas of medium vulnerability are in either the first or the second quintile.

Table 20: Comparison of IVS categories and BrazDep Quintiles in Belo Horizonte, column percentages

		IVS			
		Low	Medium	High	Very high
BrazDep	1	98.3	48.6	0.1	0.0
	2	1.6	49.6	37.7	0.0
	3	0.2	1.6	59.4	48.2
	4	0.0	0.2	2.7	49.8
	5	0.0	0.0	0.0	2.0
Total		100	100	100	100

Areas of high or very high vulnerability are overwhelmingly in the 3rd and 4th quintiles. The percentage of sectors in the 5th most deprived quintile is very small as very few areas in Belo Horizonte are in that category. To conclude, the comparison of both the continuous and the categorical versions of the two measures (IVS and BrazDep) provide very similar results in Belo Horizonte. Census sectors that have high health vulnerability are also very likely to be deprived and vice versa.

7.1.2 São Paulo

The state of São Paulo developed the first Social Vulnerability Index (*Índice Paulista de Vulnerabilidade Social*, IPVS) in 2002 with the aim of managing public services more effectively (State of São Paulo 2013). The 2010 version of the IPVS (based on the 2010 census) was released in 2013. The IPVS consists of two domains, socioeconomic and demographic, and these are provided separately for urban and rural areas. The demographic domain is strictly demographic in that it includes information on the head of household's age and the proportion of young people. The socioeconomic domain includes multiple variables on income and one variable on literacy, but does not include any information on housing conditions (e.g. electricity, sanitation, water supply) or neighbourhood.

The IPVS is a categorical measure, developed from the two domains using cluster analysis. (A single continuous vulnerability score is not provided.) Based on the cluster analysis the socioeconomic factor was divided into three and the demographic domain into two categories. The census sector type and the urban-rural classification are added to this classification resulting in seven different vulnerability groups. This results in three separate categories for high vulnerability based on the sector type – the high vulnerability urban group (type 5, including urban areas), the very high vulnerability urban group (type 6, includes subnormal urban sectors), and high vulnerability rural areas (type 7). These types are not necessarily ordered, e.g. high rural vulnerability does not mean more vulnerable compared to high or very high urban vulnerability.

The state of São Paulo comprises of 645 municipalities and 66,096 census sectors, of which the IPVS is provided for 59,773 (90%) census sectors, excluding sectors with less than 50 permanent private households. The BrazDep measure is calculated for 64,491 (98%) sectors in São Paulo, including all those for which the IPVS is provided. This means that there are 4718 sectors for which the BrazDep is provided, but the IPVS is not. Of these 26% of sectors are quite small with population below 50 people. On average, the sectors with no IPVS are

Table 21: Distribution of census sectors and population by IPVS categories and BrazDep deciles, São Paulo

IPVS	Lowest	Very low	Low	Medium	High Urban	Very high Urban	High Rural
Sectors	8.3	42.4	16.6	18.2	8.8	4.1	1.6
Population	6.1	40.1	18.0	19.2	11.1	4.4	1.0

BrazDep	Low 1	2	3	4	5	6	7	8	9	High 10
Sectors	19.4	18.1	15.6	14.2	11.4	8.8	6.8	4.3	1.2	0.1
Population	17.0	18.5	17.4	16.0	12.8	9.1	5.7	2.8	0.7	0.0

more likely to be in the above average deprived BrazDep deciles (deciles 6th to 10th, see Table 22).

Table 21 shows the distribution of the census sectors and population by both the IPVS categories and BrazDep deciles. Both show that the majority of sectors and population in São Paulo are among the low vulnerability or least deprived (deciles 1 through 3) categories. Relative to the rest of Brazil, São Paulo is not very deprived.

Since there is no single continuous IPVS score, we have compared the BrazDep measure to the socio-economic vulnerability domain. High values of the IPVS socio-economic vulnerability domain correspond to low vulnerability. To ease comparison to BrazDep, the socio-economic vulnerability domain scores were reversed by multiplying by -1. The distributions of the two measures are quite different in that BrazDep is skewed to the deprived, right-hand side of the distribution and the socio-economic vulnerability domain is skewed to the left, least vulnerable end of the distribution. As the variables used in the BrazDep measure are fairly different from the IPVS socio-economic domain we may expect the two measures to provide different results. The correlation is, however, reasonably strong at 0.68. The relationship is shown in the scatter plot – while there is a lot of scatter, the relationship is fairly obvious. The correlation of BrazDep to the demographic domain is very weak (r = -0.28).

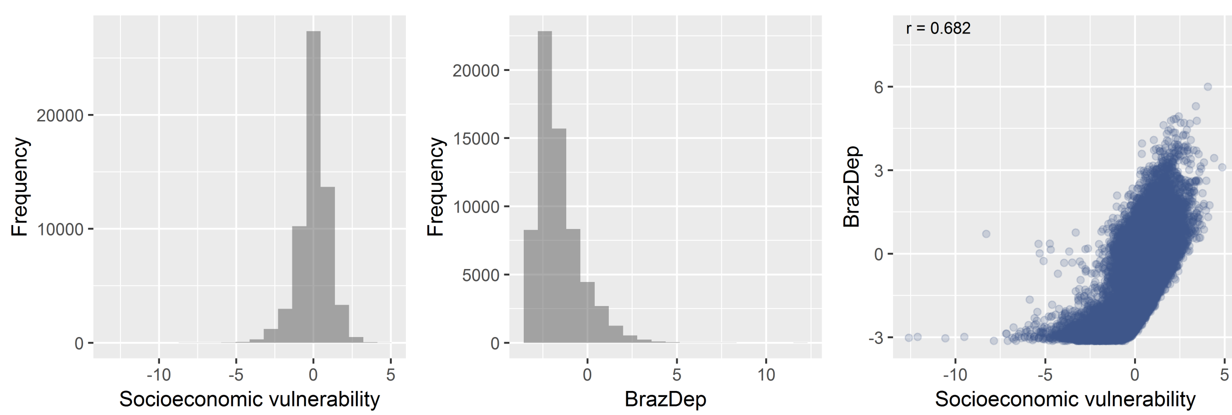


Figure 12: São Paulo census sector measures

Table 22 compares the BrazDep deciles to the IPVS categories and this also shows a good correspondence between the two measures. 80% of the least vulnerable areas are in the 1st

Table 22: Comparison of IPVS categories and BrazDep deciles in São Paulo, column percentages

BrazDep	Lowest	Very low	Low Low	Medium	High urban	Very high urban	High rural	No IPVS
1	80.2	27.7	10.3	0.0	0.0	0.0	0.0	9.9
2	12.1	35.1	17.6	0.2	0.0	0.0	0.0	7.5
3	4.0	22.5	25.6	11.9	0.5	0.0	0.0	6.4
4	1.3	7.5	23.7	33.1	14.3	2.5	0.0	8.6
5	0.8	2.0	11.3	27.9	33.9	13.9	0.2	11.5
6	0.9	1.6	5.3	16.9	28.9	28.0	2.2	14.1
7	0.5	2.4	3.9	7.3	16.1	30.5	16.1	18.0
8	0.2	1.1	2.1	2.4	5.5	20.6	46.9	16.5
9	0.0	0.0	0.1	0.3	0.8	4.2	33.8	6.1
10	0.0	0.0	0.0	0.0	0.0	0.3	0.7	1.3
Total	100	100	100	100	100	100	100	100

(least deprived) decile and a further 12% are in the second decile. The majority of sectors with very low or low vulnerability are also among the below average deprived deciles (1st to the 4th). 81% of areas with medium vulnerability are in the average deprived 4-6th deciles and areas of high or very high vulnerability are among the most deprived deciles. The vulnerable rural sectors are more likely to be among the deprived deciles compared to the vulnerable urban sectors.

Overall, the results show that the BrazDep measure is well correlated to the IPVS socio-economic domain and the BrazDep deciles and the IPVS categories provide fairly similar results. While there are some differences, both measures are likely to identify the same census sectors as either vulnerable or deprived.

7.2 Comparison to municipal level measures

There are 5565 municipalities in Brazil with population size varying from 805 in Borá to 11,253,503 in São Paulo. The distribution of municipalities by population size is shown in Figure 13. The majority of municipalities have a population size of 20,000 or fewer and about 5% have a population above 1 million. Overall, municipalities are very large and heterogeneous. Due to this, any measure of deprivation, development or vulnerability has to be interpreted with some caution, as they can mask significant inequalities and differences in deprivation or development within a municipality.

There are at least two measures similar to BrazDep that have been developed at the municipal level for the whole of Brazil. The first is the municipal Index of Social Vulnerability (*Índice de Vulnerabilidade Social*, MIVS) developed by Brazilian Institute of Applied Economy Research (IPEA) and the second is the municipal Human Development Index (MHDI) developed by a group of organizations, including the UN. While these two measures try to capture different concepts (vulnerability and development), they are similar enough to the concept of deprivation for a meaningful comparison and a potential tool for validating the deprivation measure.

Index of Social Vulnerability for Brazil. The municipal IVS, provided for the whole of Brazil, consists of three domains (and 16 indicators), each domain having the same weight on

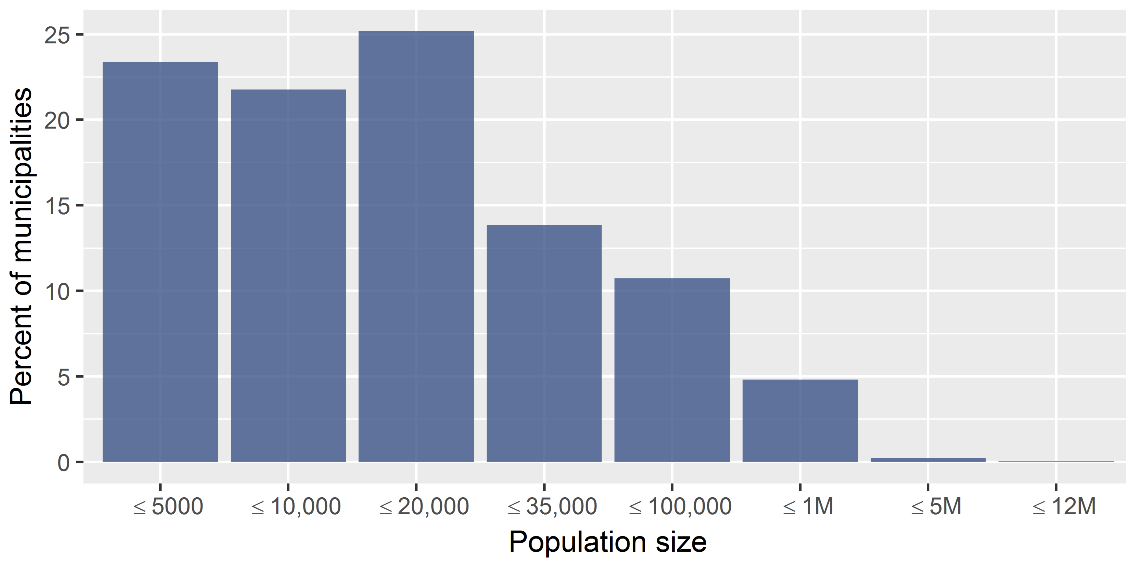


Figure 13: Municipalities by population size

the full measure (IPEA 2019).

The *human capital* domain is composed of eight indicators: 1) Infant mortality (weight: 0.125); 2) Percentage of children aged 0-5 not attending school (weight: 0.125); 3) Percentage of children aged 6-14 not attending school (weight: 0.125); 4) Percentage of women aged 10-17 with children (weight: 0.125); 5) Percentage of mothers who are heads of households, without having completed middle-school education and with at least one child under 15 years of age (weight: 0.125); 6) Percent of people not literate, aged 15 and over (weight: 0.125); 7) Percentage of children in households where none of the residents has completed elementary education (weight: 0.125); 8) Percentage of people aged 15-24 years who do not study or work and are vulnerable to poverty (weight: 0.125).

The *income and labour* domain consists of five indicators: 1) Proportion vulnerable to poverty (weight: 0.200); 2) Unemployment rate of population aged 18 and over (weight: 0.200); 3) Percentage of people 18 and over without having completed middle-school and in informal employment (weight: 0.200); 4) Percentage of people in households vulnerable to poverty and dependent on the elderly (weight: 0.200); 5) employment rate of people aged 10-14 (weight: 0.200).

Finally, the *urban infrastructure* domain is composed of three indicators: 1) Percentage of the population living in urban households without garbage collection service (weight: 0.300); 2) Percentage of people in households with inadequate water supply and sewage (weight: 0.300); 3) Percentage of people in households with *per capita* income below 1/2 minimum wage and who spend more than one hour go to work (weight: 0.400).

Municipal Human Development Index. The MHDI is a simpler measure compared to the IVS and consists of three domains – longevity, education and income – that are combined by taking the geometric mean ($MHDI = \sqrt[3]{L \times E \times I}$) (The Atlas 2019a). *Longevity* is measured simply as life expectancy at birth and the *income* domain as per capita monthly income. The *education* domain is divided into two aspects: educational level and educational flow. The education level of the adult population is measured as the percentage of people aged 18 or older with completed primary education. The educational flow of young people is an arithmetic mean of (1) the percentage of children aged 5-6 attending school; (2) the percentage of young people aged 11-13 attending the final years of primary school; (3) the percentage of young people aged 15-17 with completed primary education; and (4) the percentage of young people aged 18-20 with completed secondary education. The educational flow is given twice the weight of the education level domain score.

The three domains and the combined MHDI score vary between 0 and 1, with high values corresponding to high human development. To ease comparison to the other measures the scale of the MHDI and its domains has been reversed by subtracting the scores from 1. This leads to high values of MHDI and its domains corresponding to low development, and to positive correlations to BrazDep deprivation and IVS vulnerability indices.

The distributions of the three measures are compared in Figure 14. The municipal level BrazDep measure is created using the same z-score method (see Section 6) as for the census sector level measure. All distributions are similar and distinctly bimodal with the Southern regions having, on average, lower deprivation and vulnerability and higher development compared to the Northern regions.

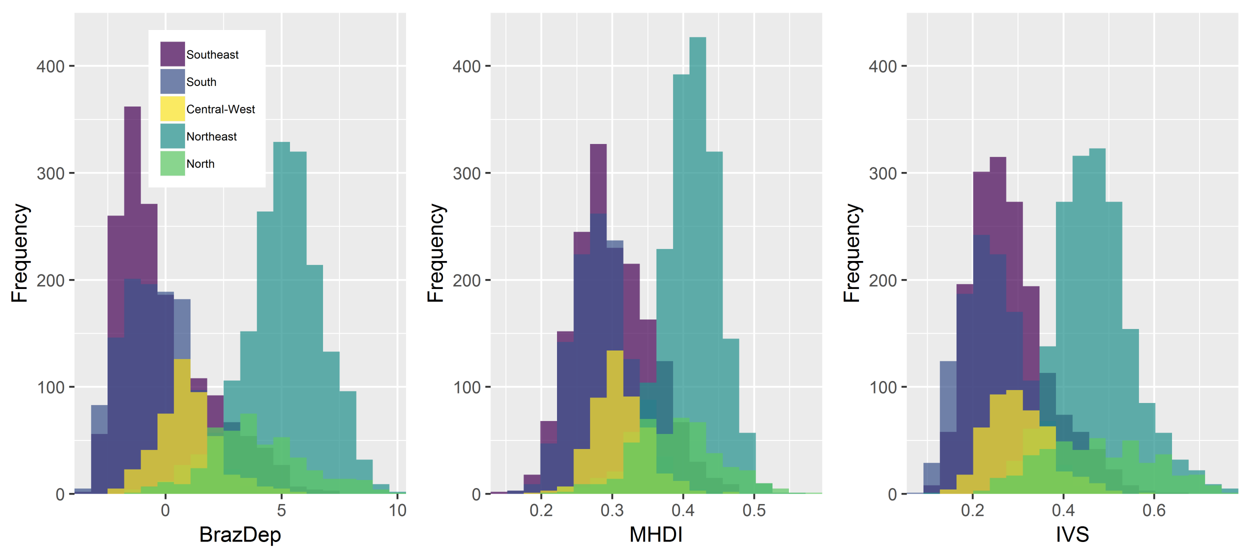


Figure 14: Distributions of BrazDep, MHDI and IVS at municipal level

The correlations between the three measures are fairly strong and are shown in Figure 15. The relationship between BrazDep and MHDI is somewhat stronger than with the IVS. Table 23 also gives the correlations of the domains from the three measures with BrazDep, IVS and MHDI. The IVS the urban infrastructure domain, based on housing variables, has a weak correlation with BrazDep and MHDI. This might be why the IVS has a slightly weaker correlation with MHDI and BrazDep. Compared to the income and education domains, the BrazDep housing domain also has a somewhat weaker correlation to IVS and MHDI. Generally, though, all of the different domains have consistently strong correlations with the

Table 23: Correlation between domains and measures, municipal indices

	BrazDep	IVS	MHDI
BrazDep: Income	0.96	0.91	0.93
BrazDep: Literacy	0.94	0.82	0.89
BrazDep: Housing	0.90	0.78	0.81
IVS: Urban infrastructure	0.58	0.83	0.55
IVS: Human capital	0.89	0.91	0.92
IVS: Inc/Employ	0.92	0.89	0.90
MHDI: Education	0.85	0.80	0.95
MHDI: Longevity	0.84	0.79	0.85
MHDI: Income and employment	0.94	0.88	0.95

three measures.

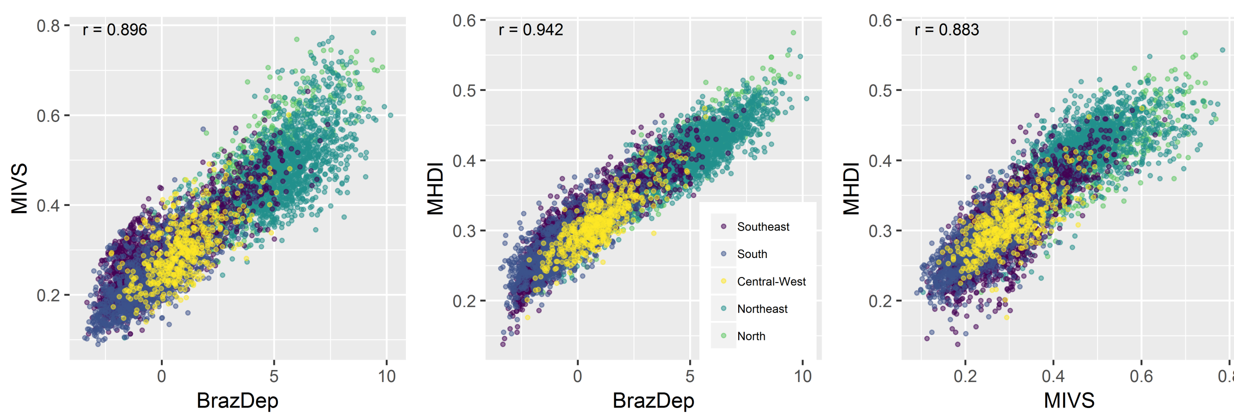


Figure 15: Relationships between BrazDep, MHDI and IVS at municipal level

We can also compare the categorical versions of the three measures to each other. Cut-off points are given for a five-category MIVS measure (IPEA 2019) and a categorical MHDI (The Atlas 2019a), both with the groups Very low, Low, Medium, High, Very high. For BrazDep we calculated population weighted quintiles. Calculating deciles and quintiles for the municipal measures is somewhat more difficult due to the large population size of some municipalities and also the variation in their size. This can lead to quintiles and deciles of fairly unequal population size. Table 24 gives the distribution of municipalities and population by the categorical deprivation, vulnerability and development measures. For BrazDep the distribution of municipalities by quintiles is very unequal. About half the municipalities are in the most deprived quintile and another 28% are in the 4th quintile. The MIVS and MHDI were not designed to have an equal proportion of people in the five categories and give a slightly different distribution.

The population distributions for the BrazDep is closer to 20% for each quintile, though the proportion of people is slightly higher in the most deprived quintile. The MHDI and the MIVS have fairly similar population distributions. About half the people are in the Very Low and Low categories of MIVS and MHDI and only about 20% in the High and Very High MIVS and 11% in High and Very high MHDI categories.

Table 24: Distribution of municipalities and population by deprivation categories, percentages

	Low 1	2	3	4	High 5
Municipalities					
BrazDep	4.0	6.1	15.3	28.1	46.6
MHDI	0.8	33.9	40.1	24.6	0.6
MIVS	11.2	30.6	22.6	21.2	14.4
Population					
BrazDep	18.4	18.2	19.9	20.1	23.4
MHDI	15.7	51.5	21.5	11.0	0.3
MIVS	10.9	41.9	27.1	12.1	8.0

Table 25: Comparison of BrazDep quintiles to IVS categories and MHDI quintiles, column percentages

		IVS categories					MHDI quintiles				
		Very Low	Low	Medium	High	Very High	Very Low	Low	Medium	High	Very High
BrazDep	1	27.2	2.9	0.1	0.0	0.0	86.4	9.5	0.1	0.0	0.0
	2	24.3	9.6	1.8	0.0	0.0	9.1	17.6	0.0	0.0	0.0
	3	37.2	29.5	9.5	0.0	0.0	4.5	40.1	4.3	0.0	0.0
	4	11.3	55.1	40.2	4.0	0.0	0.0	32.5	42.4	0.1	0.0
	5	0.0	2.9	48.4	96.0	100.0	0.0	0.3	53.2	99.9	100
Total		100	100	100	100	100	100	100	100	100	100

Table 25 compares the IVS and MHDI to the BrazDep quintiles. That BrazDep quintiles are developed differently compared to the MIVS and MHDI categories and place more areas and people in the deprived category by design. Regardless the different methods of creating the BrazDep quintiles and the MIVS categories, the two correspond reasonably well. Over half the municipalities that have very low vulnerability also have low deprivation (quintiles 1 and 2). As vulnerability increases, so the probability of being in a more deprived BrazDep quintile increases. Areas of high or very high vulnerability are overwhelmingly in the most deprived 5th quintile. The relationship between the MHDI and BrazDep quintiles is also fairly strong. Of all areas in the most developed 1st MHDI quintile 86% are in the least deprived BrazDep quintile. As development decreases, deprivation increases and 88% of the least developed municipalities are in the most deprived quintile.

Across Brazil, three different measures developed to capture deprivation, social vulnerability and development are well related to each other and tend to give similar results for the majority of municipalities. This supports the validity of the BrazDep measure.

7.3 Relationship to life expectancy, mortality and survival at municipal level

The data released with the MHDI includes five municipal level variables measuring life expectancy, mortality and survival that can be used to further validate the BrazDep measure. These include: (1) life expectancy at birth, (2) mortality under 1 year of age, (3) mortality under 5 years of age (4) probability of survival to the age of 40 and (5) probability of survival to the age of 60. We expect the BrazDep measure to be positively associated with mortality –

as deprivation increases, so should mortality. We also expect the measure to be negatively related to life expectancy and survival – as deprivation increases, life expectancy and the probability of surviving should decrease.

The correlations of the BrazDep, MHDI and IVS indices to the five health related variables are shown in Table 26. Life expectancy and mortality have a strong correlation to BrazDep and MHDI and a slightly weaker correlation to IVS. Across Brazil survival does not correlate well to any of the three measures. The results by region are somewhat different and the correlations of life expectancy and mortality to BrazDep, MHDI and IVS are weaker in all regions, especially in Central West of Brazil. The correlation of survival to BrazDep and MHDI measures is reasonably strong in Southeast and South.

The relationships are visualized in Figure 16. For mortality and life expectancy the correlations across Brazil are driven by regional differences in both deprivation and health outcomes. For survival the relationship to deprivation is more pronounced within regions rather than across regions. In the North and Northeast the probability of survival to age 60 is higher compared to the South and Southeast at the same level of deprivation. Despite relatively low levels of deprivation in the South and Southeast, other factors reduce survival in these regions.

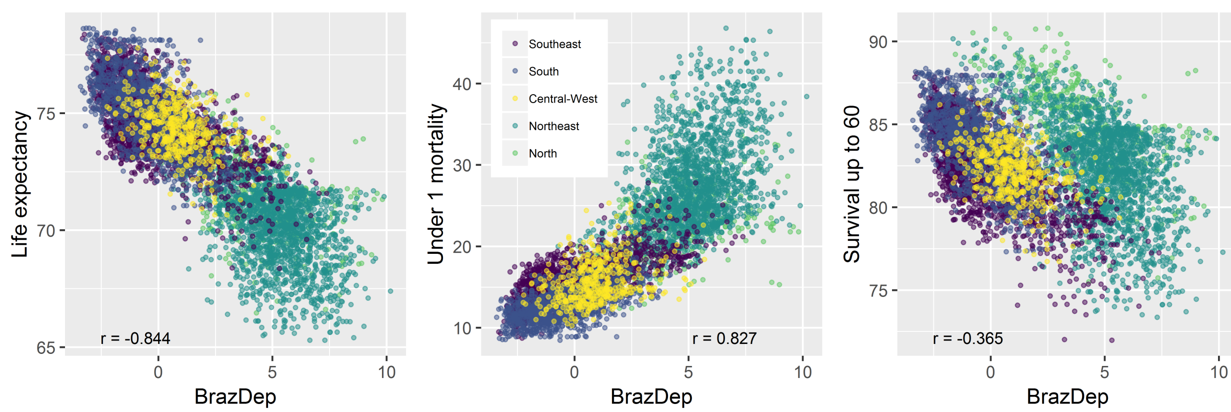


Figure 16: Correlations between BrazDep and mortality, life expectancy and survival at municipal level

Table 26: Correlation of BrazDep, IVS and MHDI to mortality, life expectancy and survival, municipal data

Health outcome	BrazDep	IVS	MHDI	MHDI, excl. life expectancy
Life expectancy	-0.84	-0.79	-0.85	-0.79
Under 1 mortality	0.83	0.78	0.83	0.77
Under 5 mortality	0.81	0.77	0.82	0.77
Survival up to 40	-0.27	-0.25	-0.35	-0.31
Survival up to 60	-0.36	-0.34	-0.45	-0.40
	Northeast			
Life expectancy	-0.50	-0.47	-0.64	-0.51
Under 1 mortality	0.50	0.51	0.63	0.51
Under 5 mortality	0.52	0.52	0.64	0.52
Survival up to 40	-0.51	-0.52	-0.64	-0.51
Survival up to 60	-0.49	-0.47	-0.63	-0.50
	North			
Life expectancy	-0.58	-0.56	-0.64	-0.56
Under 1 mortality	0.55	0.53	0.61	0.54
Under 5 mortality	0.55	0.53	0.61	0.54
Survival up to 40	-0.56	-0.53	-0.62	-0.54
Survival up to 60	-0.57	-0.55	-0.63	-0.55
	Southeast			
Life expectancy	-0.65	-0.58	-0.70	-0.62
Under 1 mortality	0.71	0.63	0.76	0.69
Under 5 mortality	0.72	0.63	0.77	0.70
Survival up to 40	-0.70	-0.58	-0.74	-0.68
Survival up to 60	-0.71	-0.62	-0.76	-0.69
	South			
Life expectancy	-0.62	-0.50	-0.65	-0.53
Under 1 mortality	0.58	0.46	0.60	0.48
Under 5 mortality	0.58	0.45	0.60	0.48
Survival up to 40	-0.63	-0.50	-0.63	-0.51
Survival up to 60	-0.62	-0.49	-0.65	-0.53
	Central-West			
Life expectancy	-0.45	-0.35	-0.51	-0.40
Under 1 mortality	0.33	0.32	0.43	0.35
Under 5 mortality	0.33	0.31	0.43	0.35
Survival up to 40	-0.45	-0.37	-0.51	-0.41
Survival up to 60	-0.44	-0.36	-0.51	-0.40

8 Uncertainty estimates

Any deprivation measure is always an estimate of "true" deprivation and, as such, there will be uncertainty about the exact level of deprivation. This uncertainty stems from multiple sources, including potential coding errors or the relative contribution of the variables included in the measure. As a result, for areas with very similar scores, it is not possible to say which is more deprived compared to others. To help us understand the extent of this uncertainty and make more reliable comparisons between areas, confidence intervals can be provided for deprivation measures.

We have used two different methods described in Brown *et al* (2014) to calculate 95% confidence intervals for the BrazDep measure. These methods consider two sources of uncertainty: the variation in the weights applied to the different variables that constitute the index, and the population size of the area of interest (a smaller population means that each of the variables has less precision). In the first of these methods the weight attached to each variable is varied and the effects of this on the score is investigated. The second approach simulates the data (i.e. the numerators for the three variables) and calculates the measure using these data.

8.1 Uncertainty in the influence of individual variables (random weights)

In the developed deprivation measure each of the three variables has been standardised and assigned equal weights. However, this does not account for situations where the distribution of the data is skewed and varies across indicators. Skewed distributions can affect the contribution of a variable to the final score. When the weights are varied the influence of some variables is reduced while the influence of others is increased. If the change in the weights, i.e. the influence of individual variables, does not affect the score much there is more confidence in the estimated level of deprivation.

To do this three random weights were drawn from a uniform distribution and then constrained to sum to three. These weights can lie anywhere from zero to three for each variable but the sum of these weights must be three for each of the random draws. The new weights were assigned to the variables and a new score calculated for all areas. This procedure was repeated 1,000 times to provide 1,000 new randomly weighted measures for all census sectors. To assess the extent of variation 2.5th and 97.5th percentiles of the generated measures were observed. These percentiles give the intervals between which 95% of the generated scores will fall, giving a likely deprivation score range for all areas.

Figure 17 show the BrazDep measure together with the 95% confidence intervals ordered by deprivation rank and grouped by deciles. The figure illustrates the general patterns in confidence intervals but all individual census sectors are not visible as each of the deciles includes about 30,000 small areas. (The handful of extreme cases omitted from the figures are included in the the tables below and in all the other analysis.) The plots show that it is often impossible to distinguish between areas within the same deprivation decile, but differences between the deciles are more obvious.

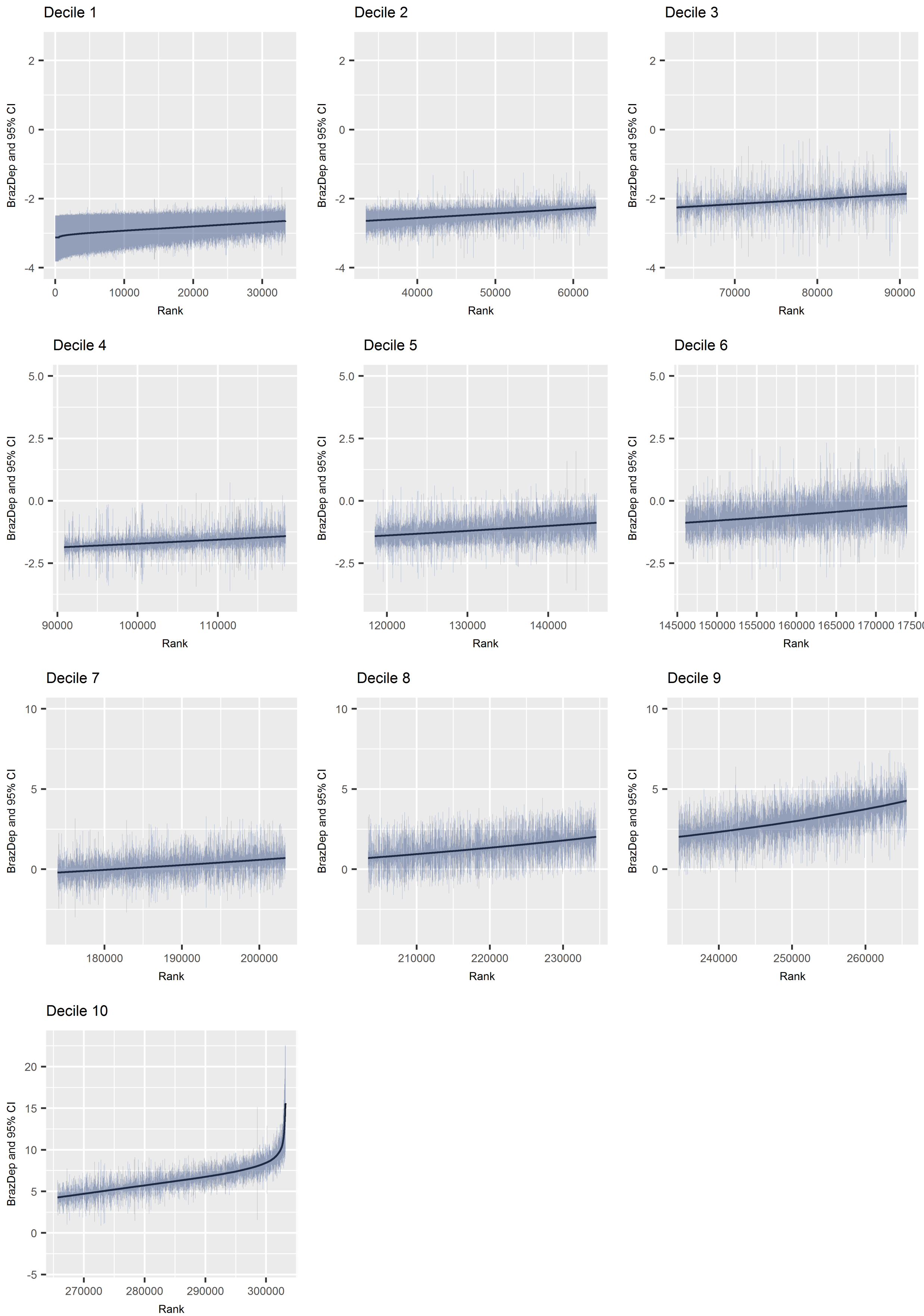


Figure 17: Uncertainty estimates, random weights

Table 27: Comparison of BrazDep deciles to the deciles from the lower and upper CIs, random weights

		BrazDep									
		Low deprivation									High deprivation
		1	2	3	4	5	6	7	8	9	10
Upper CI	1	28597	4504	0	0	0	0	0	0	0	0
	2	4549	19099	5205	0	0	0	0	0	0	0
	3	237	5471	17706	4276	0	0	0	0	0	0
	4	0	469	4294	18409	4204	0	0	0	0	0
	5	0	12	674	4405	17911	4602	0	0	0	0
	6	0	0	23	552	5127	18092	4352	0	0	0
	7	0	0	0	19	250	4811	18633	4832	0	0
	8	0	0	0	0	13	433	6347	20863	3572	0
	9	0	0	0	0	0	0	115	5428	24944	2541
	10	0	0	0	0	0	0	0	19	2610	35018
		Low deprivation									High deprivation
		1	2	3	4	5	6	7	8	9	10
Lower CI	1	29521	3245	717	322	75	35	3	16	0	0
	2	3862	20661	3788	1623	464	193	34	2	0	0
	3	0	5649	16029	4647	2148	533	212	12	1	0
	4	0	0	7365	14449	4322	1097	320	19	0	0
	5	0	0	3	6487	14442	5141	910	55	1	0
	6	0	0	0	133	5939	15177	6092	539	10	1
	7	0	0	0	0	115	5759	15878	7176	151	3
	8	0	0	0	0	0	3	5998	18792	5335	65
	9	0	0	0	0	0	0	0	4531	23136	2706
	10	0	0	0	0	0	0	0	0	2492	34784

The figure also shows that for some census sectors the confidence intervals are very large which could result in significant changes in relative rank and deprivation decile. To understand how uncertainty in estimation can affect deprivation deciles Table 27 compares the BrazDep deciles to the deciles calculated based on the lower and upper confidence intervals. To get the latter deciles the estimated deprivation score for each census sector was replaced with the upper and lower bound and the ranking was based on these values. For the majority of census sectors the decile for the measure is either the same or just one decile away from that of the CI. For the upper 95% confidence interval, 72.3% census sectors fall on the diagonal (same decile) and 26.8% are one decile away (in total 99.1% of all sectors). For the lower 95% CI there are more differences, but still 66.9% of all sectors are on the diagonal and 29.9% are one decile away from the diagonal (together 96.8% of all sectors). Accounting for both differences, 96.4% (n = 292,422) of all census sectors either have the same decile as or are one decile away from either of the CI bounds. Only 3.7% (n = 10,796) of sectors are more than one decile apart from one or both of the CI deciles. Thus, while there clearly is uncertainty about the exact numeric value of the measure, for the vast majority of the census sectors we can be fairly confident about their decile and the relative level of deprivation in comparison to other areas.

8.2 Uncertainty from small population size (random numerators)

Uncertainty in the measure can also arise from the small population size of the census sectors. The problem is universal to all small area measures and thus an appropriate statistical method (e.g. shrinkage) should be used to increase the robustness of the estimated scores. The second

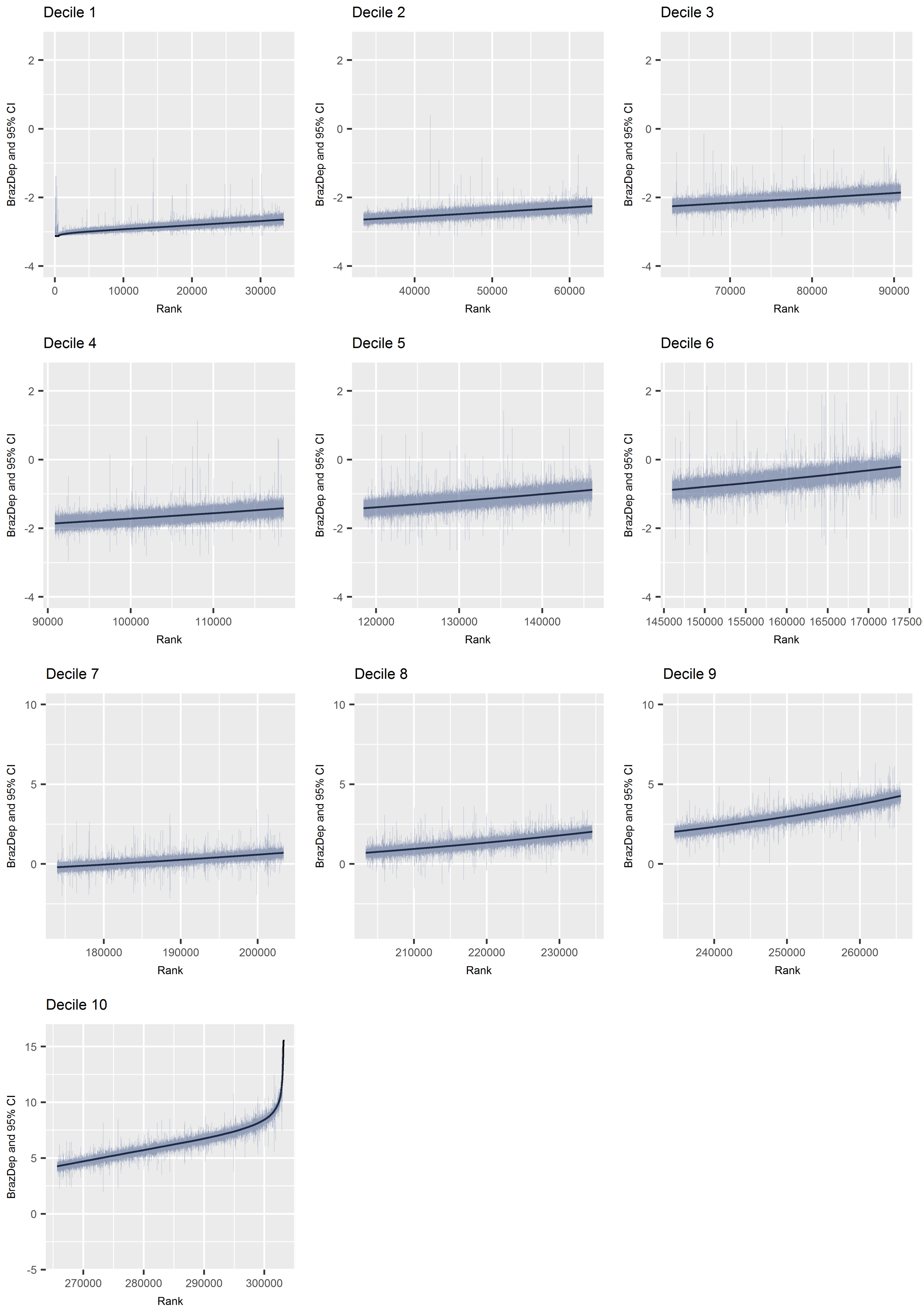


Figure 18: Uncertainty estimates, random counts

set of confidence intervals calculated here take account of uncertainty stemming from small observed numbers of cases. One thousand random samples of numerators for each of the three indicators (e.g. number of people aged 7+ who are not literate) were drawn from a binomial distribution.[4] Based on these, 1000 new deprivation measures were calculated for all census sectors. As before, the 2.5th and 97.5th percentiles of the generated measures were observed, providing the 95% confidence intervals for each area.

Table 28: Comparison of BrazDep deciles to the deciles from the lower and upper CIs, random numerators

		BrazDep									
		Low deprivation								High deprivation	
		1	2	3	4	5	6	7	8	9	10
Upper CI	1	30648	683	0	0	0	0	0	0	0	0
	2	2316	26001	798	0	0	0	0	0	0	0
	3	208	2498	24079	835	0	0	0	0	0	0
	4	108	196	2616	23496	857	0	0	0	0	0
	5	65	113	268	2877	23227	774	0	0	0	0
	6	31	46	96	301	3012	23705	683	0	0	0
	7	5	17	39	140	337	3129	24975	619	0	0
	8	2	1	6	12	70	318	3626	27158	499	0
	9	0	0	0	0	2	12	163	3342	28623	294
	10	0	0	0	0	0	0	0	23	2004	37265
		Low deprivation								High deprivation	
		1	2	3	4	5	6	7	8	9	10
Lower CI	1	32619	2776	266	97	22	9	0	6	0	5
	2	764	25912	2721	345	88	23	0	0	0	0
	3	0	867	23953	3010	402	122	22	0	0	0
	4	0	0	962	23282	3101	407	81	1	0	0
	5	0	0	0	927	23070	3232	381	43	0	0
	6	0	0	0	0	822	23431	3474	269	5	3
	7	0	0	0	0	0	714	24854	3641	111	7
	8	0	0	0	0	0	0	635	26719	2784	9
	9	0	0	0	0	0	0	0	463	27935	1890
	10	0	0	0	0	0	0	0	0	291	35645

The results from this process are shown in Figure 18. As before, the figure represents a general pattern in confidence intervals by rank and decile. It generally is difficult to distinguish sectors within the same decile, but differences are more obvious between deciles. Again, for some areas uncertainty is very large and to understand how this affects the relative level of deprivation for these census sectors we have compare the deciles calculated for the upper and lower confidence intervals to the deciles of the deprivation measure. Table 28 shows that for some areas, the change in deciles is very large and in five cases this means a shift from the most deprived to the least deprived decile. When comparing the deciles based on the upper CI to the deciles of the measure, 88.8% of sectors are on the diagonal and a further 10.4% are one decile apart (in total 99.1%). For the deciles based on the lower CI, 88.2% are on the diagonal and 10.9 are one decile apart (in total 99.1%). Taking account of differences in

[4] For each of the three variables the denominator (e.g. number of people) was used as the number of Bernoulli trials and the observed proportion (e.g. percent of people not literate) as the success rate in the sampling process. For areas where the success rate was either zero percent (e.g. there were no illiterate people) or 100% (e.g. all people were illiterate) these values were marginally changed to allow variation in the random draws. This was done by either adding 0.5 or subtracting 0.5 from the numerator before the success rate was calculated.

Table 29: Census sectors and population affected by high uncertainty, by region and urban-rural classification

Sector location	Random Weights				Random Numerator				Either			
	Sectors		Population		Sectors		Population		Sectors		Population	
	N	%	N	%	N	%	N	%	N	%	N	%
North	481	2.3	319085	2.0	55	0.3	1670	0.0	518	2.4	320354	2.0
Northeast	692	0.9	457170	0.9	174	0.2	6282	0.0	833	1.0	462440	0.9
Central-West	1552	6.6	732974	5.2	291	1.2	9934	0.1	1746	7.3	740022	5.3
South	5034	10.1	1944453	7.1	1438	2.9	51265	0.2	5907	11.4	1978252	7.2
Southeast	3037	2.3	1143219	1.4	2165	1.7	100470	0.1	4720	3.6	1223817	1.5
Rural	3312	4.7	906684	3.1	777	1.1	27742	0.1	3819	5.2	925402	3.1
Urban	7484	3.2	3690217	2.3	3346	1.4	141879	0.1	9905	4.2	3799483	2.4
Total	10796	3.6	4596901	2.4	4123	1.4	169621	0.1	13724	4.5	4724885	2.5

both the lower and upper bound of the CI, 98.6% (N = 299,095) of all census sectors either have the same decile or are one decile away from the confidence interval decile and only 1.4% (N = 4,123) of areas have bigger differences in deciles. This again shows that while there is uncertainty in the exact level of deprivation, in only a few cases is it large enough to have a substantial impact on the relative rank and decile of deprivation.

8.3 Census sectors with high uncertainty

A major benefit of calculating confidence intervals and comparing deciles for the measure and the confidence intervals is that we are able to flag and analyse areas with high uncertainty. Table 29 shows the number and percent of census sectors with high uncertainty about the deprivation level by region and urban-rural classification. High uncertainty is defined as cases where the deciles for the deprivation measure and those for any of the confidence intervals were more than one category apart. There is some regional variation in the percent of sectors with high uncertainty based on random weights. The percent of sectors with high uncertainty is highest in South (10.1%) and Central-West (6.6%) and lowest in Northeast (0.9%) of Brazil. Uncertainty measured via random weights is also slightly more frequent in rural areas (4.7%) compared to urban areas (3.2%). There is only marginal variation across regions and urban-rural classification in the percent of sectors with high uncertainty based on random numerators.

Table 29 also gives the number and percent of people in the census sectors with high uncertainty. As the sectors affected by uncertainty tend to be smaller in terms of population size, the percent of total population affected by high uncertainty in the deprivation measure is smaller than the percent of affected census sectors. In total, only about 2.5% of the population live in the sectors with high uncertainty.

To explore further how population size of a census sector is related to uncertainty, Figure 19 provides the distributions in household population by region and the type of uncertainty estimation. As expected, it clearly shows that areas with high uncertainty based on random numerators are all very small in terms of total household population. The mean population size for areas with high uncertainty using random numerators is only 37, compared to 634 for areas with low uncertainty. On average, sectors with high uncertainty based on random weights also have slightly lower population compared to areas with low uncertainty, but the differences are relatively small.

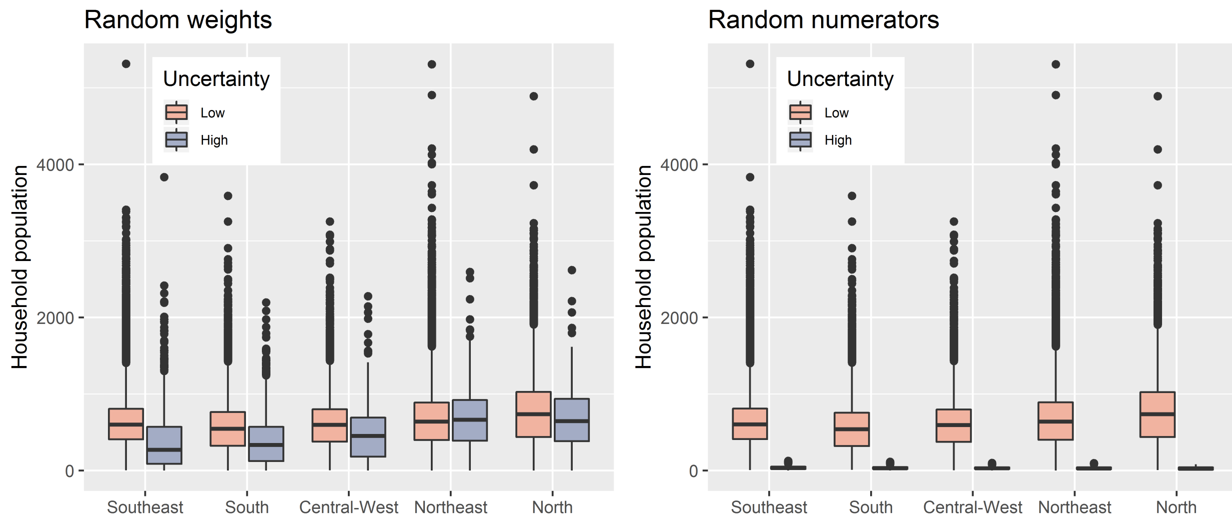


Figure 19: Population distribution by regions and level of uncertainty

Table 30 explores how large difference in the deciles for different deprivation indicators affect uncertainty. Differences in the level of deprivation estimated by the three indicators can have a substantial impact on the measure when using random weights. If one of the three indicators predicts a substantially higher level of deprivation for an area compared to the other two, then increasing its weight can radically increase the overall level of deprivation. Thus we would expect uncertainty estimated using random weights to be high for areas where the three indicators suggest very different levels of deprivation. Table 30 compares the deciles based on the housing domain to deciles based on household income and literacy for only those areas that have high uncertainty (based on random weights). The number of census sectors that are off the diagonal and have almost opposite values for different deprivation indicators is very high. Census sectors that have very different deciles for the deprivation indicators are very likely to have high uncertainty as estimated by random weights.

We have shown that there can be considerable uncertainty about the estimated level of deprivation for the census sectors. This uncertainty is caused by different factors, such as small population size and also differences in the estimated deprivation implied by the three indicators used in the measure. For a small proportion of areas (4.5% across all CS) the confidence intervals are very large and the relative rank and deprivation decile are less reliable. About 2.5% of the total population live in these census sectors. However, for the vast majority of areas (95.5% of sectors, encompassing 97.5% of the population) we can be fairly confident in the relative level of deprivation.

Table 30: Comparison of indicator deciles, areas with high uncertainty only

		Household income < 1/2 MW										Not literate, 7+									
		Low deprivation						High deprivation				Low deprivation						High deprivation			
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Poor housing	1	50	43	29	8	2	5	85	116	99	36	93	98	73	25	25	30	46	43	23	17
	2	45	19	31	7	2	0	49	77	47	25	51	46	50	24	19	19	25	48	11	9
	3	12	25	14	7	0	5	42	71	67	20	48	49	44	20	15	11	33	21	10	12
	4	14	36	19	7	0	1	31	68	60	31	32	41	45	19	5	12	40	30	16	27
	5	36	23	18	13	19	3	18	44	43	38	32	50	19	23	29	13	21	32	14	22
	6	475	82	15	61	158	9	10	18	17	22	280	150	43	76	115	119	36	21	12	15
	7	1134	998	336	24	19	12	14	13	7	8	650	620	503	332	227	125	72	22	11	3
	8	606	1050	742	63	18	12	1	9	10	10	529	515	495	386	267	206	77	32	9	5
	9	428	651	940	399	95	43	33	13	8	27	461	305	390	443	374	331	201	92	27	13
	10	71	72	168	154	85	26	11	12	12	35	105	62	71	73	76	101	87	39	18	14

9 Relationship to ethnicity/race

Some indicators of deprivation or of individual socioeconomic status are not always equally meaningful for different population groups, such as for different ages, ethnicities or race (Messer et al. 2006; Braveman et al. 2005; Krieger et al. 2003). For this reason deprivation measures should be validated for population subgroups (Allik et al. 2019). Unfortunately, we have relatively limited options for validating the BrazDep measure for different ethnic/racial groups. Currently we do not have access to health outcomes by ethnicity and thus we cannot study the effect of deprivation on health by ethnicity/race. We can, however, study the distributions and correlations of the deprivation indicators and the BrazDep measure for ethnically/racially homogeneous areas. This would indicate whether the relationship between the deprivation indicators is the same for areas of different ethnic concentration and, if so, supports the notion that the indicators have the same meaning in different sub-populations.

The census provides population data for five ethnic groups (ordered from largest to smallest): White (*branca*), Mixed (*parda*), Black (*preta*), Asian (*amarela*) and Indigenous. Approximately 90% of Brazilians belong to either of the two largest ethic groups, with 47.7% White and 43.1% of Mixed ethnicity. The other ethnic groups are quite small (7.6% Black, 1.1% Asian and 0.4% Indigenous) and for this reason the analysis here only focuses on the two biggest ethnic groups. (A small minority of 5854 people do not have a classification for ethnicity in the census.) Regionally, the proportions of different ethnic/racial groups vary (see Table 31). Almost half (48.7%) of the White population live in the Southeast and a further 23.6% in the South, but for the Mixed ethnic group these percentages are 34.9% and 5.5%. In general, people of Mixed ethnicity/race are more likely to live in the two Northern regions and those of White ethnicity in the two Southern regions. The White population is also somewhat more likely to live in urban areas compared to those of Mixed background.

As the spatial patterning of ethnic groups presented in Table 31 overlaps somewhat with the patterning of deprivation, we can expect our deprivation measure to be associated with ethnicity/race. This is indeed the case: a higher percent of White people in a census sector is correlated with lower deprivation (r = -0.61), while a higher percent of Mixed people is correlated with higher deprivation (r = 0.58). These associations are based on aggregate data for small areas and do not mean that individuals of Mixed backgrounds are always more deprived compared to Whites.

Table 31: Ethnicity by region and urban-rural classification, column percentages

Sector location	All	White	Mixed	Black	Asian	Indigenous	Unclassified
Northeast	27.8	17.2	38.4	34.9	30.3	25.4	16.9
North	8.3	4.1	12.9	7.3	8.3	37.4	4.7
Central-West	7.4	6.5	8.4	6.5	9.8	15.9	7.1
Southeast	42.1	48.7	34.9	43.8	42.7	12.1	67.1
South	14.3	23.6	5.5	7.6	8.8	9.2	4.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Rural	15.6	11.9	19.6	14.3	13.5	61.3	9.9
Urban	84.4	88.1	80.4	85.7	86.5	38.7	90.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 32: Number and percent of census sectors by predominant ethnicity*

Frequency	White	Mixed	Asian	Black	Indigenous
Number of CS	65807	18632	5	40	1576
Percent from all CS	21.7	6.1	0.0	0.0	0.5
Distribution of sectors by region					
North	0.0	29.5	0.0	15.0	57.6
Northeast	0.6	57.5	40.0	67.5	18.3
Central-West	1.3	3.2	0.0	5.0	14.7
South	51.7	0.2	0.0	2.5	6.4
Southeast	46.3	9.5	60.0	10.0	3.0
Total	100	100	100	100	100

*Predominant ethnicity is defined as 75% or more people within a census sector belonging to a single ethnic group.

To identify ethnically/racially homogeneous areas we calculated the percent of people in each of the ethnic groups for all census sectors. Table 32 shows the number and percent of census sectors where 75% or more people belong to any of the ethnic groups. Only 28% (86,060) of all sectors have a "predominant ethnicity" as defined by the 75% cut-off. The White population is predominant in 21.7% of sectors and the Mixed population in 6.1% of sectors. The regional variation is also shown in the table. Of all the predominantly White sectors, 52% are in the South and 46% in the Southeast. Of the predominantly Mixed areas, 30% are in the North and 58% in the Northeast. There are very few sectors where any of the other ethnic/racial groups make up 75% or more of the population. We also considered other cut-off points, but increasing it would reduce the number of sectors with any predominant ethnicity and reducing it would increase heterogeneity.

Table 33 shows the correlations between the deprivation indicators and the measure for the predominantly White and Mixed census sectors. The predominantly Indigenous, Asian or Black ethnicities grouped together for the correlations. In general, the correlations are quite similar and strong for predominantly White and Mixed areas. The main difference is that correlation of household income to the BrazDep measure and to the other two indicators is slightly weaker in the predominantly Mixed areas compared to the White areas. Overall though, there is little difference in the relationships between these indicators and measures between the two groups of census sectors, suggesting a consistent relationship across the two main ethnic groups.

The correlation coefficients between the three indicators are weaker in predominantly Indigenous, Asian or Black areas (r between 0.33-0.40). The correlations between the indicators and the measure are stronger ($r > 0.57$), especially between literacy and the measure ($r = 0.94$). The predominantly Indigenous, Asian and Black census sectors have very high levels of income and housing deprivation with only marginal variation between areas for these two indicators. However, there is considerable variation in literacy rates. For this reason the overall deprivation level is most strongly associated to literacy rates across the predominantly Indigenous, Asian or Black areas.

Table 33: Correlations between indicators and BrazDep, by ethnicity/race

Variable	BrazDep	Income	Literacy	Housing
Predominantly White				
BrazDep	1.00			
Household income < 1/2 MW	0.87	1.00		
Not literate, 7+	0.82	0.67	1.00	
Not adequate housing	0.88	0.61	0.58	1.00
Predominantly Mixed				
BrazDep	1.00			
Household income < 1/2 MW	0.78	1.00		
Not literate, 7+	0.87	0.55	1.00	
Not adequate housing	0.87	0.59	0.59	1.00
Predominantly Indigenous, Asian or Black				
BrazDep	1.00			
Household income < 1/2 MW	0.57	1.00		
Not literate, 7+	0.94	0.37	1.00	
Not adequate housing	0.61	0.40	0.33	1.00

10 Conclusions

Small-area deprivation measures are a popular tool for analysing the spatial patterning of deprivation and socioeconomic inequalities in health. They are used both by researchers and policy makers in many high-income countries and have also become more common in low and middle income countries. BrazDep is the first deprivation measure provided at the smallest geographical area level for the whole of Brazil. It is based on the 2010 Brazilian Population Census and places census sectors on a scale from the least to the most deprived areas.

The BrazDep measure is the result of a thorough analysis of the available census data. The indicators used to construct the deprivation measure were chosen from over 20 different variables based on their conceptual fit with deprivation, empirical variation and correlation with each other. The variables were selected to reflect the different domains of wealth, education and housing and include: (1) percent of households with per capita income below 1/2 minimum wage; (2) percent of people not literate, aged 7+; and (3) average of percent of people with inadequate access to sewage, water, garbage collection and no toilet. Three alternative methods of combining variables into a measure were compared and in the final analysis the z-score method was preferred as a simple and easily replicated approach. There differences between this and the other methods (arithmetic mean and principal component analysis) were small and had minimal impact on the results.

To further support the measure we also validated it by comparing it to health and social vulnerability measures developed at the census sector level for the municipality of Belo Horizonte and the state of São Paulo. At the municipal level we compared the BrazDep measure to the social vulnerability and human development indices across the whole of Brazil. Finally, the deprivation measure was also correlated with health indicators at the municipal level. The results from all of the above analysis supported the validity of the developed deprivation measure.

Our measure is provided together with confidence intervals and a flag for areas on high uncertainty that should help researchers and policy-makers understand the level of uncertainty in the measured deprivation. For a small proportion of census sectors uncertainty in the level of deprivation is very large and the relative rank and deprivation decile are unreliable. However, for the vast majority of sectors we can be fairly confident in the relative level of deprivation.

We hope that in the future the BrazDep deprivation measure can be used in research and policy to estimate socioeconomic inequalities in health, assess progress in achieving the Sustainable Development Goals, monitor the health of social welfare program beneficiaries (such as for *Bolsa Família* and *Minha Casa, Minha Vida*) and for many other research projects and policy programmes.

11 Appendix A. Correlations between indicators

Table A.1: Correlation between income variables

Variable	Household income		Head of household income						Personal income								
	$\leq 1/4$	$\leq 1/2$	$\leq 1/2$	≤ 1	Male $\leq 1/2$	Male ≤ 1	Female $\leq 1/2$	Female ≤ 1	No Income	$\leq 1/2$	≤ 1	Male No Inc	Male $\leq 1/2$	Male ≤ 1	Female No Inc	Female $\leq 1/2$	Female ≤ 1
Household income																	
$\leq 1/4$	1.00	0.92	0.88	0.84	0.91	0.86	0.67	0.68	0.78	0.93	0.51	0.85	0.94	0.52	0.50	0.81	0.46
$\leq 1/2$	0.92	1.00	0.81	0.93	0.80	0.92	0.68	0.82	0.78	0.95	0.69	0.79	0.89	0.70	0.56	0.89	0.61
Head of household income																	
$\leq 1/2$	0.88	0.81	1.00	0.80	0.92	0.76	0.81	0.67	0.72	0.85	0.40	0.79	0.86	0.39	0.44	0.72	0.38
≤ 1 MW	0.84	0.93	0.80	1.00	0.76	0.97	0.66	0.88	0.65	0.85	0.81	0.69	0.82	0.81	0.43	0.79	0.74
Male $\leq 1/2$	0.91	0.80	0.92	0.76	1.00	0.80	0.60	0.56	0.68	0.85	0.42	0.82	0.91	0.40	0.35	0.68	0.42
Male ≤ 1	0.86	0.92	0.76	0.97	0.80	1.00	0.57	0.78	0.62	0.86	0.81	0.70	0.84	0.82	0.39	0.78	0.73
Female $\leq 1/2$	0.67	0.68	0.81	0.66	0.60	0.57	1.00	0.71	0.66	0.69	0.31	0.58	0.60	0.34	0.57	0.71	0.25
Female ≤ 1	0.68	0.82	0.67	0.88	0.56	0.78	0.71	1.00	0.63	0.75	0.71	0.57	0.64	0.69	0.54	0.78	0.67
Personal income																	
No income	0.78	0.78	0.72	0.65	0.68	0.62	0.66	0.63	1.00	0.87	0.18	0.89	0.79	0.22	0.84	0.84	0.13
$\leq 1/2$	0.93	0.95	0.85	0.85	0.85	0.86	0.69	0.75	0.87	1.00	0.53	0.87	0.95	0.55	0.63	0.93	0.49
≤ 1	0.51	0.69	0.40	0.81	0.42	0.81	0.31	0.71	0.18	0.53	1.00	0.27	0.50	0.96	0.06	0.52	0.95
Male, no	0.85	0.79	0.79	0.69	0.82	0.70	0.58	0.57	0.89	0.87	0.27	1.00	0.92	0.23	0.51	0.69	0.30
Male $\leq 1/2$	0.94	0.89	0.86	0.82	0.91	0.84	0.60	0.64	0.79	0.95	0.50	0.92	1.00	0.48	0.42	0.76	0.49
Male ≤ 1	0.52	0.70	0.39	0.81	0.40	0.82	0.34	0.69	0.22	0.55	0.96	0.23	0.48	1.00	0.18	0.57	0.83
Female, no	0.50	0.56	0.44	0.43	0.35	0.39	0.57	0.54	0.84	0.63	0.06	0.51	0.42	0.18	1.00	0.79	-0.08
Female $\leq 1/2$	0.81	0.89	0.72	0.79	0.68	0.78	0.71	0.78	0.84	0.93	0.52	0.69	0.76	0.57	0.79	1.00	0.42
Female ≤ 1	0.46	0.61	0.38	0.74	0.42	0.73	0.25	0.67	0.13	0.49	0.95	0.30	0.49	0.83	-0.08	0.42	1.00

Table A.2: Correlation between housing variables

Variable	Network water	Network or well	Toilet or bath	Toilet and bath	Network sewage	Network or septic tank	Garbage collected	Electricity	Network electricity
Housing conditions, extended	0.76	0.77	0.64	0.77	0.73	0.84	0.89	0.47	0.47
Housing conditions, narrow	0.87	0.70	0.57	0.69	0.83	0.78	0.89	0.43	0.43
Access to network water	1.00	0.71	0.41	0.48	0.60	0.56	0.74	0.35	0.35
Access to water from network or well	0.71	1.00	0.54	0.60	0.44	0.46	0.63	0.41	0.43
Toilet or bath/shower	0.41	0.54	1.00	0.77	0.32	0.38	0.56	0.54	0.47
Toilet and bath/shower	0.48	0.60	0.77	1.00	0.41	0.48	0.65	0.56	0.55
Access to network sewage	0.60	0.44	0.32	0.41	1.00	0.83	0.57	0.23	0.24
Access to network sewage or septic tank	0.56	0.46	0.38	0.48	0.83	1.00	0.62	0.28	0.28
Garbage collected	0.74	0.63	0.56	0.65	0.57	0.62	1.00	0.42	0.41
Access to electricity	0.35	0.41	0.54	0.56	0.23	0.28	0.42	1.00	0.84
Access to network electricity	0.35	0.43	0.47	0.55	0.24	0.28	0.41	0.84	1.00

Table A.3: Correlations for neighbourhood condition variables

Variable	No Streetlight	No pavement	No sidewalk	No storm drain
No Streetlight	1.00			
No pavement	0.41	1.00		
No sidewalk	0.36	0.72	1.00	
No storm drain	0.22	0.42	0.37	1.00
Average environment	0.52	0.85	0.85	0.73
Average environment, excluding lighting	0.40	0.85	0.85	0.75

References

Allik, Mirjam. 2019. *SocEpi: Tools for for Health Inequality Analysis*. https://github.com/m-allik/SocEpi.

Allik, Mirjam, D. Brown, R. Dundas, and Alastair H. Leyland. 2016. "Developing a New Small-Area Measure of Deprivation Using 2001 and 2011 Census Data from Scotland." *Health & Place* 39 (May): 122–30. http://www.sciencedirect.com/science/article/pii/S1353829216300156.

Allik, Mirjam, Alastair Leyland, Maria Yury Travassos Ichihara, and Ruth Dundas. 2019. "Creating Small-Area Deprivation Indices: A Guide for Stages and Options." *Journal of Epidemiology & Community Health*. https://doi.org/10.1136/jech-2019-213255.

Bonfim, Cristine, Ana Maria Aguiar-Santos, Dinilson Jr Pedroza, Tadeu Rodrigues Costa, Jose Luiz Portugal, Conceicao Oliveira, and Zulma Medeiros. 2009. "Social Deprivation Index and Lymphatic Filariasis: A Tool for Mapping Urban Areas at Risk in Northeastern Brazil." *International Health* 1 (1): 78–84.

Braveman, Paula A, Catherine Cubbin, Susan Egerter, Sekai Chideya, Kristen S Marchi, Marilyn Metzler, and Samuel Posner. 2005. "Socioeconomic Status in Health Research: One Size Does Not Fit All." *Jama* 294 (22): 2879–88.

Brown, Denise, Mirjam Allik, Ruth Dundas, and Alastair H. Leyland. 2014. "Carstairs Scores for Scottish Postcode Sectors, Datazones & Output Areas from the 2011 Census - Report." MRC/CSO Social; Public Health Sciences Unit, University of Glasgow.

Cabrera-Barona, Pablo, Thomas Murphy, Stefan Kienberger, and Thomas Blaschke. 2015. "A Multi-Criteria Spatial Deprivation Index to Support Health Inequality Analyses." *International Journal of Health Geographics* 14 (March): 11.

Carr-Hill, Roy, and Paul Chalmers-Dixon. 2005. *The Public Health Observatory Handbook of Health Inequalities Measurement*. Edited by Jennifer Lin. Centre for Health Economics, University of York.

Carstairs, Vera, and Russell Morris. 1991. *Deprivation and Health in Scotland*. Aberdeen: Aberdeen University Press.

Decreto-Lei No. 311, de 02.03. 1938. "Dispõe Sobre a Divisão Territorial Do País, E Dá Outras Providências." In. Coleção de Leis do Brasil - 1938, Página 438 Vol. 1. https://www2.camara.leg.br/legin/fed/declei/1930-1939/decreto-lei-311-2-marco-1938-351501-publicacaooriginal-1-pe.html%20(accessed%2001.10.2019).

Fukuda, Yoshiharu, Keiko Nakamura, and Takehito Takano. 2007. "Higher Mortality in Areas of Lower Socioeconomic Position Measured by a Single Index of Deprivation in Japan." *Public Health* 121 (3): 163–73. http://www.sciencedirect.com/science/article/pii/S003335060600326X.

Green, Mark A., Konstantinos Daras, Alec Davies, Ben Barr, and Alex Singleton. 2018. "Developing an Openly Accessible Multi-Dimensional Small Area Index of 'Access to Healthy

Assets and Hazards' for Great Britain, 2016." *Health & Place* 54 (November): 11–19. http://www.sciencedirect.com/science/article/pii/S1353829218303563.

Havard, Sabrina, Séverine Deguen, Julie Bodin, Karine Louis, Olivier Laurent, and Denis Bard. 2008. "A Small-Area Index of Socioeconomic Deprivation to Capture Health Inequalities in France." *Social Science & Medicine* 67 (12): 2007–16. https://doi.org/http://dx.doi.org/10.1016/j.socscimed.2008.09.031.

IBGE, Instituto Brasileiro de Geografia e Estatística. 2010. "Censo 2010. Summary of Survey Steps." 2010.

———. 2011. "Base de Informações Do Censo Demográfico 2010: Resultados Do Universo Por Setor Censitário." Rio de Janeiro. 2011.

Ichihara, Maria Yury Travassos, Dandara Ramos, Poliana Rebouças, Flávia Jôse Oliveira, Andrêa J. F. Ferreira, Camila Teixeira, Mirjam Allik, et al. 2018. "Area Deprivation Measures Used in Brazil: A Scoping Review." *Revista de Saude Publica* 52:83. https://doi.org/https://doi.org/10.11606/S1518-8787.2018052000933.

IPEA, Instituto de Pesquisa Econômica Aplicada. 2019. "Atlas Da Vulnerabilidade Social." June 2019. http://ivs.ipea.gov.br/index.php/pt.

Jarman, B. 1983. "Identification of Underprivileged Ares." *Br Med J* 286: 1705–8.

Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, David H. Rehkopf, and S. V. Subramanian. 2003. "Race/Ethnicity, Gender, and Monitoring Socioeconomic Gradients in Health: A Comparison of Area-Based Socioeconomic Measures:The Public Health Disparities Geocoding Project." *Am J Public Health* 93 (10): 1655–71. https://doi.org/10.2105/AJPH.93.10.1655.

Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson. 2002. "Geocoding and Monitoring of Us Socioeconomic Inequalities in Mortality and Cancer Incidence: Does the Choice of Area-Based Measure and Geographic Level Matter?: The Public Health Disparities Geocoding Project." *American Journal of Epidemiology* 156 (5): 471–82.

Lalloué, Benoît, Jean-Marie Monnez, Cindy Padilla, Wahida Kihal, Nolwenn Le Meur, Denis Zmirou-Navier, and Séverine Deguen. 2013. "A Statistical Procedure to Create a Neighborhood Socioeconomic Index for Health Inequalities Analysis." *International Journal for Equity in Health* 12 (1): 21. https://doi.org/10.1186/1475-9276-12-21.

Leyland, Alastair H., Ruth Dundas, Philip McLoone, and F. Andrew Boddy. 2007. "Cause-Specific Inequalities in Mortality in Scotland: Two Decades of Change. A Population-Based Study." *BMC Public Health* 7 (1): 172. https://doi.org/10.1186/1471-2458-7-172.

Lòpez-De Fede, Ana, John E. Stewart, James W. Hardin, and Kathy Mayfield-Smith. 2016. "Comparison of Small-Area Deprivation Measures as Predictors of Chronic Disease Burden in a Low-Income Population." *International Journal for Equity in Health* 15 (May): 89. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4901405/.

Martin, David, Philip Brigham, Paul Roderick, Sarah Barnett, and Ian Diamond. 2000. "The (Mis) Representation of Rural Deprivation." *Environment and Planning A* 32 (4): 735–52.

McConnachie, Alex, and Chris Weir. 2005. "Evaluation of Statistical Techniques in the Scottish Index of Multiple Deprivation." University of Glasgow. https://www2.gov.scot/simd2004evaluation.

Messer, Lynne C, Barbara A Laraia, Jay S Kaufman, Janet Eyster, Claudia Holzman, Jennifer Culhane, Irma Elo, Jessica G Burke, and Patricia O'Campo. 2006. "The Development of a Standardized Neighborhood Deprivation Index." *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 83 (6): 1041–62. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3261293/.

Morris, Russell, and Vera Carstairs. 1991. "Which Deprivation? A Comparison of Selected Deprivation Indexes." *Journal of Public Health* 13 (4): 318–26.

NHS Health Scotland. 2019. "Keep Well." July 2019. http://www.healthscotland.com/keep-well.aspx.

Noble, Michael, Helen Barnes, Gemma Wright, and Benjamin Roberts. 2009. "Small Area Indices of Multiple Deprivation in South Africa." *Social Indicators Research* 95 (2): 281. https://doi.org/10.1007/s11205-009-9460-7.

Noble, Michael, Gemma Wright, George Smith, and Chris Dibben. 2006. "Measuring Multiple Deprivation at the Small-Area Level." *Environ Plan A* 38 (1): 169–85. https://doi.org/10.1068/a37168.

Norman, Paul, Laurie Berrie, and Daniel Exeter. 2019. "Calculating a Deprivation Index Using Census Data." *Australian Population Studies* 3 (1): 30–39. http://www.australianpopulationstudies.org/index.php/aps/article/view/45.

Pampalon, R., D. Hamel, P. Gamache, A. Simpson, and M. D. Philibert. 2014. "Validation of a Deprivation Index for Public Health: A Complex Exercise Illustrated by the Quebec Index." *Chronic Diseases and Injuries in Canada* 34 (1): 12–22.

Pink, Brian. 2013. "Socio-Economic Indexes for Areas (Seifa)." Australian Bureau of Statistics.

Prefeitura Belo Horizonte. 2013. "Indice de Vulnerabilidade Da Saude 2012." Prefeitura Belo Horizonte. https://prefeitura.pbh.gov.br/sites/default/files/estrutura-de-governo/saude/2018/publicacaoes-da-vigilancia-em-saude/indice_vulnerabilidade2012.pdf.

Revelle, William. 2018. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. https://CRAN.R-project.org/package=psych.

Salmond, Clare E, and Peter Crampton. 2012. "Development of New Zealand's Deprivation Index (Nzdep) and Its Uptake as a National Policy Tool." *Canadian Journal of Public Health* 103.

Santos, Maria de Lourdes Sperli Geraldes, Silvia Helena Figueiredo Vendramini, Claudia Eli Gazetta, Sonia Aparecida Cruz Oliveira, and Tereza Cristina Scatena Villa. 2007. "Poverty: Socioeconomic Characterization at Tuberculosis." *Revista Latino-Americana de Enfermagem* 15 Spec No: 762–7.

Sánchez-Cantalejo, Carmen, Ricardo Ocana-Riola, and Alberto Fernández-Ajuria. 2008. "Deprivation Index for Small Areas in Spain." *Social Indicators Research* 89 (2): 259–73. https://doi.org/10.1007/s11205-007-9114-6.

Smith, Tom, Michael Noble, Stefan Noble, Gemma Wright, David McLennan, and Emma Plunkett. 2015. "The English Indices of Deprivation 2015." Department of Communities; Local Government.

State of São Paulo. 2013. "Indice Paulista de Vulnerabilidade Social." http://www.iprs.seade.gov.br/ipvs2010/view/.

The Atlas, Atlas of Human Development in Brazil. 2019a. "Atlas of Human Development in Brazil." United Nations Development Programme UNDP; Institute of Applied Economic Research IPEA; João Pinheiro Foundation FJP. June 2019. http://www.atlasbrasil.org.br/2013/en/o_atlas/idhm/.

———. 2019b. "Methodology. Human Development Units." United Nations Development Programme UNDP; Institute of Applied Economic Research IPEA; João Pinheiro Foundation FJP. October 2019. http://atlasbrasil.org.br/2013/en/o_atlas/metodologia/construcao-das-unidades-de-desenvolvimento-humano/.

The Scottish Government. 2012. "Scottish Index of Multiple Deprivation 2012." A National Statistics Publication for Scotland.

Townsend, Peter. 1987. "Deprivation." *Journal of Social Policy* 16 (02): 125–46. https://doi.org/10.1017/S0047279400020341.

Townsend, Peter, P. Phillimore, and A. Beattie. 1988. *Health and Deprivation: Inequality and the North*. Croom Helm.

UNDP, United Nations Development Programme. 2018. "Human Development Indices and Indicators: 2018 Statistical Update.technical Notes." United Nations. http://hdr.undp.org/en/content/human-development-index-hdi.

Vasquez, Baltica AND Tunstall, Andrea AND Cabieses. 2016. "Where Are Socioeconomically Deprived Immigrants Located in Chile? A Spatial Analysis of Census Data Using an Index of Multiple Deprivation from the Last Three Decades (1992-2012)." *PLOS ONE* 11 (1): 1–19. https://doi.org/10.1371/journal.pone.0146047.