UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Department of Social Statistics and Demography

Volume 1 of 1

Evaluating the accuracy and feasibility of gridded population sampling to overcome bias due to missing populations in household surveys

by

Dana Renee Thomson-Browne

Thesis for the degree of Doctor of Philosophy

March 2020

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

In low- and middle-income countries (LMICs), household survey data are a main source of information for planning, evaluation, and decision-making. LMIC survey methods have not changed substantially in forty years, while population trends and technologies have. This mismatch is likely excluding an increasing number of vulnerable and mobile populations and leading to erosion in LMIC household survey data accuracy. When populations are under-represented in surveys, they are less likely to benefit from funds allocated by governments and international agencies to promote health and social equity. This thesis evaluates the accuracy of using gridded population sample frames as an alternative to census sample frames, and the feasibility of implementing gridded population surveys in a real-world setting.

To compare accuracy of census and gridded population sample frames, a realistic “true” population was simulated in Khomas, Namibia for 2016. The simulated population was derived from Namibia’s 2011 census and 2013 Demographic and Health Survey comprising several household outcome indicators and latitude-longitude coordinates digitised from actual building locations in satellite imagery. Sixteen versions of realistic outdated-inaccurate censuses were simulated by (a) removing households in buildings not present in satellite imagery in 2011, 2006, and 2001, and (b) randomly removing a percentage of rural, urban slum, and urban non-slum households based on a literature review of LMIC census undercounts. For each realistic census, a gridded population sample frame was modelled using WorldPop methodologies. Accuracy was then assessed by drawing repeated samples from the simulated censuses and gridded population datasets, and comparing them to the “true” population. To evaluate feasibility, a gridded population survey was conducted in Kathmandu, Nepal and focus group discussions and observation were made with survey implementers.

In the accuracy assessment, I found that the main source of error in gridded population datasets was not outdated, inaccurate census inputs, but rather, the WorldPop practice of using average population density from aggregated census units to predict population density in finer grid cells. Despite inaccuracies, I found that WorldPop methods disaggregated population in or around settlements, and that the distribution of population estimates in grid cells within urban areas were +/-20% the “true” population distribution. Gridded population sample frames derived from outdated or inaccurate census data produced more accurate survey estimates than the original outdated, inaccurate census EAs frames. In the Kathmandu feasibility assessment, gridded population sampling proved to be cost-effective and feasible, and use of an area-microcensus design identified more vulnerable and mobile households than a comparable two-stage sample.

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Social Statistics and Demography

Thesis for the degree of Doctor of Philosophy

EVALUATING THE ACCURACY AND FEASIBILITY OF GRIDDED POPULATION SAMPLING TO OVERCOME BIAS DUE TO MISSING POPULATIONS IN HOUSEHOLD SURVEYS

Dana Renee Thomson-Browne

Table of Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, Dana Renee Thomson-Browne, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Evaluating the accuracy and feasibility of gridded population sampling to overcome bias due to missing populations in household surveys

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:
   - Thomson DR, Stevens FR, Ruktanonchai NW, et al. (2017)
   - Thomson DR, Kools L, Jochem WC. (2018)
   - Thomson DR, Bhattarai R, Khanal S, et al. (2019)
   - Thomson DR, Kuffer M, Boo G, et al. (2020)
   - Thomson DR, Rhoda DA, Tatem AJ, et al. (2020)

Signed: ....

Date: ....

Acknowledgements

I would like to thank my supervisors Dr. Andrew Tatem, Dr. Nikos Tzavidis, Dr. Tomas Bird, and Dr. Doug Leasure for their support and guidance through this work. I would also like to thank reviewers Dr. David Martin, Dr. Angela Luna, Dr. Graham Moon, Dr. Patrick Sturgis, and Dr. Ryan Engstrom for their helpful and detailed feedback on early concept notes and drafts of this thesis.

The evaluation of gridded population sampling in a real-world setting was made possible by the Surveys for Urban Equity Team coordinated by University of Leeds in the UK, HERD International in Nepal, ARK Foundation and CIPRB in Bangladesh, and Hanoi University of Public Health in Vietnam. I would like to thank everyone on this team, in particular Dr. Helen Elsey, Dr. Ak Narayan Poudel, Dr. Joseph Hicks and Mr. Christopher Wright at University of Leeds, and Mr. Radheshyam Bhattarai, Mr. Rajeev Dhungel, Mr. Subash Gajurel, Ms. Sudeepa Khanal, Ms. Shraddha Manandhar, and Dr. Sushil Baral at HERD International.

Numerous survey experts, software programmers, and survey practitioners have provided essential feedback over several years that enabled development of the GridSample R package, GridSample.org, and gridded population survey field protocols. I wish I could thank each person individually, but there are too many people to name. Among the survey experts is Dale Rhoda at BioStat Global, Felicity Cutts at London School of Hygiene and Tropical Medicine, Mashreky Saidur at CIPRB in Bangladesh, Dominique Habimana at National Institute of Statistics in Rwanda, and Chris Jochem at University of Southampton. Huge thanks to Forrest Stevens at University of Louisville and Nick Rucktanonchai at University of Southampton who programmed the GridSample R package. Also, enormous thanks to Ian Waldock at University of Southampton who programmed GridSample.org. Among the gridded population survey practitioners, I would like to thank Sergio Regi and Laure Boudinaud at the World Food Programme, Jacquee Bunnell Hesse who contracts with World Vision International, Michael Imohi and staff at the National Bureau of Statistics in Nigeria, and Yuliya Dudaronak and colleagues at ORB International.

Substantial computing resources were required for this work, and I am grateful for use of the IRIDIS High Performance Computing Facility at the University of Southampton. I also thank the South Coast Doctoral Training Programme and the UK Economic and Social Research Council for funding my study.

Finally, and very sincerely, I thank my husband, friends, and family for providing feedback on thesis drafts, and being generous with their patience and support as I journeyed through the PhD process.

Definitions and Abbreviations

AIS	AIDS Indicator Survey
Ambient	Ambient population is the 24-hour average of day- and night-time population
ArcGIS	A platform to create, manage, analyse, and map spatial data
ARK	Advancement through Research and Knowledge
CDC	Centers for Disease Control and Prevention
CDR	Call Detail Record
CIPRB	Centre for Injury Prevention and Research Bangladesh
CPS	Contraceptive Prevalence Survey
CRAN	Comprehensive R Archive Network
CV	Coefficient of Variation
DALYs	Disability Adjusted Life Years
DHS	Demographic and Health Survey
EA	Enumeration Area
EPI	Expanded Programme on Immunization
FGD	Focus Group Discussion
FP-MCH	Family Planning-Maternal Child Health
GeoODK	A tablet-based questionnaire tool that can visualize and store spatial location data
GHDx	Global Health Data Exchange
GHSL	Global Human Settlement Layer
GHS-POP	Global Human Settlement - Population layer
GIS	Geographic Information System
GitHub	An open-source software development platform
GPS	Global Positioning System
GPW	Gridded Population of the World
gridEZ	An algorithm to create gridded enumeration zones from gridded population data
GRUMP	Global Rural-Urban Mapping Project
GUF	Global Urban Footprint
HERD	(formally) Health Research and Social Development Forum
HIC	High Income Country
HRSL	High Resolution Settlement Layer
ICC	Inter-cluster correlation
iD Editor	An online, user-friendly tool to edit OpenStreetMap
IDP	Internally Displaced Person

LDC	Least Developed Country
LMIC	Low- or Middle-Income Country
LSMS	Living Standards Measurement Survey
MAE	Mean Absolute Error
MAUP	Modifiable Areal Unit Problem
MICS	Multiple Indicator Cluster Survey
MIS	Malaria Indicator Survey
MSE	Mean Square Error
NSA	National Statistical Agency
OSM	OpenStreetMap, a free crowd-sourced map of the world
OSMAnd	A tablet-based application for field maps and navigation with OpenStreetMap data
PES	Post Enumeration Survey
PPES	Probability Proportionate to Estimated Size
PPS	Probability Proportionate to Size
PSU	Primary Sampling Unit
PubMed	A database of peer-reviewed scientific literature from health and medical fields
R	A language and environment for statistical computing and graphics
RHS	Reproductive Health Survey
RMSE	Root Mean Square Error
RTI	(formerly) Research Triangle Institute
Scopus	A database of peer-reviewed scientific literature from physical and social sciences
SDG	Sustainable Development Goals
Stata	A statistical software package
STEPS	STEPwise Approach to Surveillance Surveys
SSU	Secondary Sampling Unit
SUE	Surveys for Urban Equity
UNEP	United Nations Environment Programme
UNICEF	United Nations Children’s Emergency Fund
USAID	United States Agency for International Development
VAM	Vulnerability Analysis and Mapping
WFS	World Fertility Survey
WHO	World Health Organization
WPE	World Population Estimate

Chapter 1: Introduction

In low- and middle-income countries (LMICs), household health and economic survey methods and tools have not changed substantially in forty years. However, in the same time period, population trends, as well as the methods and tools to measure populations, have changed significantly. Innovations in survey methods and tools have been widely adopted in high-income countries, for example, collection of survey responses via phone and the internet. However, in LMICs, the continued use of outdated survey methods and tools in rapidly changing societies has excluded important sub-populations and led to an erosion in survey data accuracy.

Household health and economic surveys are essential in this era of the Sustainable Development Goals (SDGs) where all countries aim to achieve 17 goals by 2030, including zero extreme poverty and zero hunger whilst leaving no one behind (UN-DESA, 2018). In LMICs, household surveys are often the only source of information about dozens of the 232 indicators used to track progress toward the SDGs (IAEG-SDGs, 2017). In countries where other population data exist, surveys may still be used because survey results are easily disaggregated by income, sex, age, race, and other characteristics which are required for SDG reporting (Diaz *et al.*, 2017). SDGs are measured for all countries, but are especially important in LMICs where poverty and poor health are highly prevalent, and often unequally distributed (Diaz *et al.*, 2017).

SDG and other development indicators are important for planning at all scales – city/district, national, and international (UN-DESA, 2014; UN-Habitat, 2016). Decision-makers involved with city planning or district development require representative, timely data about multiple population characteristics, disaggregated by sub-group and local area to design appropriate policies, programmes, and infrastructure (UN-Habitat, 2016). Regional and national policy-makers depend on the same data to prioritize agendas and budgets, and to make decisions. International decision-makers, namely funders, also use these data, and need them to be comparable across countries to prioritize global agendas and to target funding (IAEG-SDGs, 2017). Household surveys are thus, additionally, attractive to collect standardized indicators across countries and over time.

A large number of household surveys are performed on an *ad-hoc* basis, and often focus on a single topic. However, with the financial support of routine survey programmes, multi-topic nationally-representative surveys implemented by national statistical agencies have become common place. Routine survey programmes such as the Demographic and Health Surveys (DHS) (ICF International, 2017), Multiple Indicator Cluster Surveys (MICS) (UNICEF, 2019), and Living Standard Measurement Surveys (LSMS) (Grosh and Munoz, 1996) were each established in the 1980s to collect hundreds of standardized health, demographic, social, and economic indicators about populations in LMICs. These surveys use essentially the same methods and tools, which are considered the gold-standard among development practitioners.

Together the DHS, MICS, and LSMS comprise a large portion of the data being used to monitor the SDGs and to make development decisions. It is standard practice to report SDG indicators disaggregated by gender, wealth, geographic area, and other sub-groups. Not only does disaggregation reveal socioeconomic disparities in health and economic outcomes, it can suggest issues of under-representation in the underlying household survey data. For example, fewer urban poor households are sampled in many routine household surveys than would be expected (Elsey *et al.*, 2016).

In this thesis, I review evidence that millions of people are unintentionally excluded from the DHS, MICS, LSMS, and other similar surveys worldwide. Exclusion happens first, in census sample frames, second, during survey design, and third, during survey implementation. Additional populations, including homeless, nomads, prisoners and people living in care-homes, are intentionally excluded from surveys by design because they are not considered to be part of the general, residential target population. Intentionally and unintentionally excluded populations in household surveys are overwhelmingly poor, vulnerable, and/or mobile.

When populations are excluded or under-represented in surveys, they are much less likely to benefit from the billions of pounds in funding allocated each year by governments and international aid agencies to promote health and social equity in LMICs (Development Initiatives, 2018). If survey coverage of the general residential population improves, we can expect a larger portion of respondents to be poor, vulnerable, and/or mobile. Quantifying the size of intentionally excluded populations could further contextualize our measurements of the general residential population for decision-makers. With this more accurate information, governments and donors stand to make greater, and more equitable, impacts on the health and livelihoods of people in LMICs.

1.1 Outline of thesis

In Chapter 2, I present evidence that populations in LMICs have undergone massive change over the last four decades, particularly in urban areas, and that household survey methods have not changed since the 1980s. I argue that outdated survey methods and tools, likely under-represents the most vulnerable populations in standard household survey data, leading to bias, misclassification and poorly informed decisions about the populations which are often of greatest interest in survey data.

In Chapter 3, I outline requirements for a simulated, realistic population. Here I summarize the methods and parameters developed in collaboration with colleagues to generate a realistic, synthetic population from a census sample, household survey, spatial covariates, and digitized building locations. The second part of this chapter includes a systematic literature review and summary of types of populations that are under-counted in LMIC censuses, including ranges of under-counts in rural, urban slum, and urban non-slum populations. I also estimate change in population over time (2001-2016) based on satellite imagery. I use these findings to generate 16 realistic census scenarios in which population are missing at low, middle, and high rates, and the data are five, ten, and 15 years out of date.

Chapter 4 is a study of survey sample frame accuracy. This chapter reviews the methods and parameters that I used to model each of the simulated census scenarios as a gridded population dataset using WorldPop's Random Forest model and publicly available spatial covariate datasets. This chapter ends with an analysis of gridded population dataset accuracy at the scale of grid cells, and evaluates whether accuracy of gridded population estimates can be improved by aggregating cells.

Chapter 5 is a study of gridded population survey accuracy in a typical LMIC urban context. I open this chapter with a review of standard survey methods and a review of existing gridded population surveys, comparing the ways in which gridded population sampling and census-based sampling are similar or different. The focus of this chapter is an assessment of gridded population survey accuracy, which I evaluate with repeated samples from a census sample frame and two gridded population sample frames under various scenarios of outdated, inaccurate input data.

Chapter 6 is a study of survey implementation methods and tools, specifically the feasibility of implementing gridded population surveys in complex urban environments. This chapter describes a real-world implementation of a gridded population survey in Kathmandu, Nepal, a highly complex city where standard surveys are thought to omit the urban poorest (Elsey *et al.*, 2016). In this gridded population survey, I randomize sampling areas to either an area-microcensus or two-stage sample design, allowing me to compare types of households that are sampled in an area-microcensus survey but are missed using a typical two-stage survey design. Next I describe qualitative data collection methods and results from the implementation, including the experience of survey planners and survey implementers.

1.2 Contributions

1.2.1 Contribution to state of knowledge

- **Quantification of populations who are unintentionally excluded or misclassified** in census-based household surveys.
- Understanding of the specific **methods and protocols that lead to unintentional exclusion and misclassification** of vulnerable and mobile populations in census-based household surveys.

1.2.2 Methodological contribution

- **Requirements for simulation of a realistic population** geo-located to realistic building locations in a LMIC context.
- **A method to assess cell-level accuracy** of gridded population datasets.
- **A method to quantify the impact of outdated, inaccurate census data** on the accuracy of top-down gridded population datasets.
- **A method to compare the accuracy of gridded population surveys and typical census-based surveys** with regard to estimated population indicators.

1.2.3 Contribution to practice

- Perform the **first cell-level accuracy assessment** of gridded population data
- Influence the **types of accuracy assessments** performed by gridded population data modellers, moving the field from accuracy assessments as the scale of the input data (e.g., administrative units) to accuracy assessments at the scale of the output data (i.e., grid cells).
- **Feasible methods to implement gridded population surveys** and improve representation of households and individuals who might otherwise be excluded from standard household surveys.
- Influence **more equitable targeting of development resources** to vulnerable and mobile populations in LMICs.

1.3 Publications

During the course of this thesis, the following publications stemmed directly from this work:

Thomson DR, Stevens FR, Ruktanonchai NW, Tatem AJ, Castro MC. 2017. GridSample: An R Package to Generate Household Survey Primary Sampling Units (PSUs) from Gridded Population Data. Int J Health Geogr; 16: doi: 10.1186/s12942-017-0098-4.

Thomson DR, Kools L, Jochem WC. 2018. Linking synthetic populations to household geolocations: a demonstration in Namibia. Data; 3(3): doi: 10.3390/data3030030.

Thomson DR, Bhattarai R, Khanal S, Manandhar S, Dhungel R, Gajurel S, et al. 2019. Addressing unintentional exclusion of vulnerable and mobile households in traditional surveys in Kathmandu, Dhaka, and Hanoi: A mixed-methods feasibility study. Journal of Urban Health; doi: 10.1007/s11524-020-00485-z.

Thomson DR, Kuffer M, Boo G, Hati B, Grippa T, Elsey H, et al. Need for an Integrated Deprived Area "Slum" Mapping System (IDEAMAPS) in LMICs. 2020. Social Sci; 9(5): doi: 10.3390/socsci9050080.

Thomson DR, Rhoda DA, Tatem AJ, Castro MC. Gridded population survey sampling: A systematic scoping review of the field and strategic research agenda. 2020. Int J Health Geogr; 19: doi: 10.1186/s12942-020-00230-4.

During the course of this thesis, the following publications were related to this work:

Elsey H, Poudel AN, Ensor T, Mirzoev T, Newell JN, Hicks JP, et al. 2018. Improving household surveys and use of data to address health inequities in three Asian cities: protocol for the Surveys for Urban Equity (SUE) mixed methods and feasibility study. BMJ Open; 8(11): doi:10.1136/bmjopen-2018-024182.

Thomson DR, Linard C, Vanhuysse S, Steele JE, Shimoni M, Siri J, et al. 2019. Extending data for urban health decision-making: A menu of new and potential neighbourhood-level health determinants datasets in LMICs. Journal of Urban Health; 96(4): doi: 10.1007/s11524-019-00363-3.

Ensor T, Bhattarai R, Manandhar S, Poudel AN, Dhungel R, Baral S, et al. From Rags to Riches: Assessing poverty and vulnerability in urban Nepal. 2020. PLoS One; 15(2): doi: 10.1371/journal.pone.0226646.

Amer S, Thomson DR, Chew R, Rose A. Worldwide population estimates for small geographic areas: can we do a better job? In: Big Data Meets Survey Science. 2020. Wiley: Hoboken NJ.

Chapter 2: Motivation

2.1 Personal motivation

I planned and implemented my first household survey as a graduate student at Harvard School of Public Health in 2010. One of the two *mwamis* (traditional kings and administrators) on Idjwi Island, DR Congo, sought to influence the dozens of international organizations located in eastern DR Congo to invest in the island's health system. Idjwi is located in the middle of Lake Kivu, which separates DR Congo from Rwanda. By 2010, Idjwi had long been a destination for hundreds of thousands of people escaping violence in Rwanda, Burundi, Uganda, and DR Congo (IRIN, 1996). Further, population growth had been essentially unconstrained for decades due to lack of availability of family planning services, along with personal preferences for large families to ensure security in a context of high child mortality (Romanuik, 2011).

The *mwami* sought to build new facilities, supply basic medicines, and improve staff training so that the sick no longer had to choose between no meaningful care, or use of prohibitively expensive and infrequent motorboats to Goma or Bukavo in mainland DR Congo for care. The *mwami* contacted Jacques Sebisaho, a clinician and former resident of Idjwi living in the US, to organize data collection about the health situation on Idjwi. The aim was to both provide the *mwami* with evidence to present to potential donors, and to brandish this evidence as reputable by inviting researchers from an international recognized institution to collect it. Dr. Sebisaho reached out to the Harvard Humanitarian Initiative, and organized a multi-disciplinary team of Harvard graduate students from public health, medicine, policy, and design to conduct several studies on Idjwi during the summer of 2010. I developed the methods and design for a multi-topic household survey, and Michael Hadley oversaw the survey fieldwork and data cleaning. Together we analysed and reported results of the survey to the *mwami*, Idjwi officials, and organizations located in eastern DR Congo.

I was initially flummoxed about how to select a representative sample of households on Idjwi. Then, and now, DR Congo's last census was collected in 1984. Not only were population estimates for Idjwi grossly outdated, they were not sufficiently disaggregated to serve as a household survey sample frame. I had trained and worked as a geographer for the Demographic and Health Survey programme, and was aware of the LandScan gridded population estimates produced in 1 kilometre by 1 kilometre grid cells for the globe. Out of desperation, I decided to use LandScan as a sample frame, though at the time, no other survey team teams had reported use of a gridded population sample frame. Instead, other household surveys in DR Congo had either performed time-consuming and expensive enumeration activities to update the census sample frame (MdP and Macro International, 2008), or listed villages with population estimates from local leaders (Coghlan *et al.*, 2006). As students with limited funding, we did not have the capacity or time to use either of these approaches.

With support of my supervisor, Dr. Marcia Castro, and faculty members Dr. Bethany Hedt-Gauthier and Dr. Gunther Fink, I developed a new approach to select survey clusters of approximately 40 households each with probability proportionate to population size, and derived sample probability weights (Thomson *et al.*, 2012). All survey preparations, including selection of 50 sampling units and production of field maps, were performed in the US before we first visited Idjwi. I was nervous whether navigation with GPS devices and maps of printed satellite imagery from Google Earth would be feasible and sensible for the field teams. I was also nervous that the gridded population estimates might be wrong.

However, to our relief and to the benefit of the survey results, the field teams found the maps of printed satellite imagery and overlaid cluster boundaries intuitive to use. Many had never seen a laptop computer before, let alone satellite imagery, but these “pictures from the sky” showed familiar paths, landmarks (e.g., large trees) and buildings clearly. Further, where LandScan estimated higher population density, we observed greater density of buildings in satellite imagery, and successfully located and interviewed those structures on the ground. I have since been involved with a number of typical census-based household surveys, and appreciate, now, that the Idjwi survey fieldwork ran as smoothly as any survey team could wish for.

During preparations for the Idjwi survey, I received an email announcing the release of a new open-source gridded population dataset called AfriPop, which had more disaggregated estimates than LandScan. Following the Idjwi survey, I contacted the AfriPop team lead, Dr. Andrew Tatem, and described the potential high-impact use of gridded population data as a survey sample frame in settings with outdated, inaccurate censuses. He agreed to volunteer the time of a researcher and programmer on his team, Dr. Forest Stevens, to create an algorithm to automate gridded population sampling methods. Over the next couple of years, Dr. Stevens and I worked remotely to develop and release the GridSample R package, and presented this work at an international Health Systems Research conference in 2014.

Through this conference, I met Dr. Helen Elsey, an urban health researcher at University of Leeds, and Dr. Sushil Baral, director of a public health research organization in Kathmandu, Nepal. Dr. Elsey, Dr. Baral, and their research teams were struggling at the time to explain why the urban poorest in Kathmandu appeared to be grossly underrepresented in typical household survey data, even when the surveys were implemented by highly experienced teams using gold-standard protocols. They wondered if the typical household survey methodology itself might unintentionally exclude the urban poorest, and whether a new sample frame, and/or new survey protocols could more accurately represent the urban poorest in surveys. In 2014 and 2015, I worked with Dr. Elsey, Dr. Baral and others to outline possible ways in which poor urban households might be unintentionally excluded in typical census-based LMIC household surveys.

We set out to test our hypotheses with our own gridded population survey in Kathmandu in 2015, though it was interrupted by devastating earthquakes in Nepal that year. We had completed the mapping and listing of households in nearly all of our 90 survey clusters by the time the first earthquake struck, which provided some evidence about household and dwelling configurations in the city. This experience and our limited data helped us to further flesh out our hypotheses about how and why urban poor households are underrepresented in standard surveys (Elsey *et al.*, 2016).

Following surveys in both Idjwi and Kathmandu, I became deeply concerned that huge numbers of the most vulnerable people in the poorest countries on this planet, including rural-to-urban migrants and urban slum dwellers, remain invisible in global health statistics. I was further concerned that the best efforts by governments and donors to meet the needs of the poorest, and growing urban inequities, are doomed to fail without accurate survey data. I pursued this thesis with Dr. Tatem and colleagues to evaluate tools and methods that might improve the accuracy of household survey data such that the decisions and resources allocated to development in LMICs stand a chance of improving the lives of the most vulnerable.

2.2 Routinely collected population data

In this section, I review available sources of routinely collected population demographic, health and economic data, and consider the extent to which these data sources are timely, representative, and disaggregated for use by planning and development decision-makers at city/district, national, and international levels. The main sources of routinely collected population data are decennial censuses, administrative records, and household surveys. The focus of this section is on population counts, which are useful for decision-making and for selecting representative household surveys, as well as population characteristics such as health or wealth status.

2.2.1 Decennial censuses

Censuses are a complete count of the population at a point in time conducted by a national government. A typical census requires billions of pounds and several years of preparation, is conducted over the course of a few weeks, and includes age, sex, education, and occupation information for each person, as well as their household membership, and household physical structure information. A count of people in the general residential population who do not respond should be imputed by the responsible national government agency (UNSD, 2010). Institutional populations, such as those living in army barracks or care-homes, as well as nomadic populations and homeless street-sleeping populations, are counted separately, and should be adjusted for under-counts using post-enumeration techniques and imputation (UNSD, 2010; Skinner, 2018). Depending on the country, informally settled populations, such as people living in unregistered slums and non-citizens, are sometimes intentionally excluded (UNSD, 2010). Censuses are customarily conducted decennially (every 10 years), though many LMICs do not perform censuses on schedule due to funding shortfalls, civil conflict, or political crisis.

The quality and frequency of decennial censuses have steadily improved in LMICs over time. However, in Africa, Asia, and Latin America during the 2010 round, census accuracy was questionable in 13 countries that failed to correct for missing, invalid or inconsistent responses in their censuses (UNSD, 2010). Without corrections, some of the poorest and most vulnerable people in these countries would have been excluded from the census, and a majority of those excluded (e.g., street sleepers, informally settled populations, institutional populations, and non-citizens) would have been from cities (Carr-Hill, 2013). Seventeen additional countries in these regions failed to administer a census on schedule (UNSD, 2019). Given the speed and magnitude of urbanization in African and Asian countries, even high-quality census data can become outdated within a few years of collection. These issues are explored in greater depth in Chapter 3.

While well-conducted censuses are representative of the population and can be disaggregated to local area and by socio-demographic characteristics, the data are extremely expensive, collected infrequently, and cover few indicators used in planning and development initiatives such as the Sustainable Development Goals (SDGs). In a LMIC national statistical agency (NSA), paying for a well-conducted census comes at the expense of other quality data. The 2010 Ghana Population and Housing census, for example, cost 72 million US dollars and accounted for 90% of the NSA total budget between 2010 and 2013 (Chen *et al.*, 2013). Censuses serve as an ideal survey sample frame, if they are corrected for missing responses and are not too outdated, though both of these conditions are rarely met in LMICs.

2.2.2 Administrative records

Administrative records include birth and death registries, health and crime databases, and tax and land records, which together can provide near-real-time information about population counts and their demographic, health, and economic conditions. While administrative records can be timely and often disaggregated to a local level, they are considered to be egregiously biased in LMICs due to incomplete coverage and/or poor quality of administrative record-keeping (Mahapatra *et al.*, 2007; Setel *et al.*, 2007). Part of the problem is that administrative data are extremely expensive to collect and maintain, as they depend on the entire government apparatus working at full capacity, continuously.

Only a few LMICs have been able to improve their most basic vital registration systems over the last 50 years. For example, an estimated 50 million new born children still go unregistered each year worldwide (Mahapatra *et al.*, 2007). Of the 75 countries in Africa and Asia that reported vital registration data to the UN for the 2015 Demographic Yearbook, only 9 (28%) African countries and 26 (60%) Asian countries were thought to have registered more than 90% of their national births (UNSD, 2016). Administrative records are thus not considered a population-representative data source on their own; they are used cautiously, or not at all, in decision-making, and never serve directly as a survey sample frame.

2.2.3 Representative household surveys

In the absence of representative administrative data and infrequent censuses in LMICs, governments tend to turn to nationally-representative household surveys for data. Household surveys cover far more indicators, and are substantially more affordable than either censuses or administrative records. The use of nationally-representative probability survey samples originated in wealthy countries. In the United States (US), for example, the first representative probability surveys were used to assess the labour force during The Depression and World War II (Groves *et al.*, 2009). These methods evolved in the 1940s and 1950s with consumer surveys, the 1960s with crime and victimisation surveys, and the 1970s with the first population-level health surveys (Groves *et al.*, 2009). By 1980, the US government and international agencies began funding large-scale routine health surveys, both domestically and internationally.

The Global Health Database Exchange (GHDx) registered 1,585 nationally-representative multi-topic household probability surveys in 177 countries between 1980 and 2014[1]. These surveys focus on *reproductive and child health* – 39 Reproductive Health Surveys (RHS), 297 Demographic and Health Surveys (DHS), and 236 Multiple Indicator Cluster Surveys (MICS); *chronic illness* – 159 STEPwise Approach to Surveillance Surveys (STEPS); *malaria or HIV/AIDS* – 22 Malaria Indicator Surveys (MIS) and 10 AIDS Indicator Surveys (AIS); *livelihoods* – 119 Living Standard Measurement

[1] I use January 1, 2015 as a cut off because the database was not complete for recent surveys which were being cleaned and published at the time of this writing in 2017.

Surveys (LSMS); and at least 703 other national household surveys on similar topics (IHME, 2020). Over 100 countries have had five such surveys since 1980, and 62 countries have had ten or more such surveys in the same time frame (IHME, 2020).

As statistical methods and computer-based databases have advanced, governments and international donors have increased the capacity to administer and use survey data in LMICs, resulting in a steady increase in the number of national household surveys implemented by government over time (Figure 1). Although some argue that these investments have come at the expense of strengthening administrative data collection systems (Setel *et al.*, 2007), the reality is that survey data are the main source of information used for planning and development decision-making in LMICs today. Survey data are often preferred to census and administrative data because they are (intended to be) representative of national and sub-national populations, can be disaggregated by sub-group, collected more frequently than census data, use standardized questionnaires across countries, cover hundreds of indicators each, and cost much less to implement (Murray, 2007).

The private sector collects dozens of additional nationally-representative surveys globally to monitor consumer trends (Neilsen, 2017) and public opinions (Gallup, 2017a). Rapid surveys are also conducted after natural disasters and during conflicts to assess and respond to urgent needs (HDX, 2017; WFP, 2020).

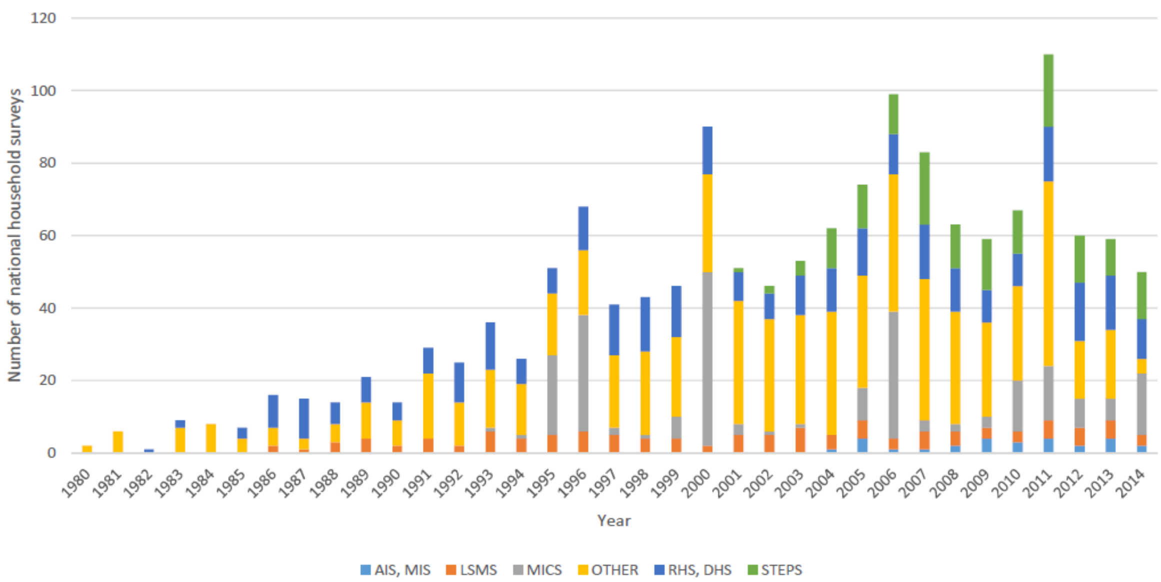


Figure 1. Number of national health and economic household surveys in LMICs, 1980-2014

Data from GHDx database (IHME, 2020)

The focus of this thesis will be on routine national household probability surveys implemented by national governments as part of the DHS, MICS, and LSMS programmes because the results of these surveys are often adapted as official national statistics and widely used for development decision-making. However, given the similarity of methods and tools, implementers of rapid assessment, such as the Vulnerable Analysis and Mapping (VAM) surveys by World Food Programme (WFP, 2020), and private sector surveys, for example, the World Poll by Gallup (Gallup, 2017a), will find the issues, methods, and results in this thesis relevant to their surveys.

2.3 Standard census sampling

In this section, I review standard census-based sampling practices used by DHS, MICS, LSMS, and other routine surveys. These are the currently accepted gold-standard survey methods to which I will compare gridded population sampling in this thesis. Key manuals describing standard survey design and implementation include: DHS Survey Sampling and Household Listing Manual (ICF International, 2012a), MICS 5 Manual for Mapping and Household Listing (UNICEF, 2013), and the LSMS Manual for Planning and Implementing the Living Standards Measurement Study Survey (Grosh and Munoz, 1996).

Routine nationally-representative health and economic household surveys are generally comprised of 10,000 or more households, cost approximately half a million US dollars each, and take roughly two years to plan, implement and publish (UNSD, 2005b). As part of trends toward decentralized government and increased disaggregation of health and economic statistics to address local disparities, several surveys have been much larger and more expensive including the 2014 Kenya DHS which included over 36,000 households and was representative of 46 separate counties (KNBS *et al.*, 2015), and the 2015-16 India National Family Health Survey (DHS equivalent) with over 600,000 households (IIPS and ICF International, 2017).

Small areas and then households are sampled at random from a complete dataset of the population, usually a census. I reviewed the survey designs of all 157 Standard and Interim DHSs conducted between 2000 and 2016 to get a sense of contemporary survey design practices (Table 1) (ICF International, 2020). Of the reviewed surveys, 94.3% used the last census as a sample frame, and the remaining 5.7% used some other official government listing of areas or households. Further, 51.0% used a sample frame more than five years old, and 12.7% used a sample frame more than 10 years old, with the average sample frame age being 6.6 years. Given that DHS methods are widely replicated; I consider these results to be reflective of broader survey practice.

Table 1. Sample frame source and age, Demographic and Health Surveys 2000-2016

Data from Demographic and Health Surveys (ICF International, 2020)

Region	Number of surveys 2000-2016	Census sample frame Percent	Other government sample frame Percent	Age of sample frame in years Mean [Range]	Sample frame >5 years old Percent	Sample frame >10 years old Percent
Africa	90	95.6	4.4	6.8 [0-30]	52.2	12.2
Asia, Oceania	39	87.2	12.8	5.5 [1-15]	41.0	5.1
Central & South America	19	100.0	0.0	8.2 [0-18]	63.2	36.8
Europe, Middle East	9	100.0	0.0	5.7 [1-9]	55.6	0.0
All	157	94.3	5.7	6.6 [0-30]	51.0	12.7

Complex sampling techniques are used to achieve the greatest amount of statistical power while minimizing costs. These techniques include stratification and multistage cluster sampling. Stratification means that independent samples are drawn from each sub-region of the country (Figure 2). Two-stage cluster design means that census enumeration areas (EAs) with non-zero populations are sampled first with probability proportionate to population size (PPS), and then households are listed in selected EAs and sampled (UNSD, 2005b). Three-stage designs are sometime used where census EAs are not available or when the EA listing is extremely long (UNSD, 2005a). In a three-stage design, larger administrative areas are sampled first with PPS, smaller administrative areas are sampled second at random or with PPS, and households are listed and sampled third. Example three-stage cluster surveys include the 2005 Egypt DHS (El-Zanaty and Way, 2006) and 2015 Colombia DHS (MINSALUD and Profamilia, 2015).

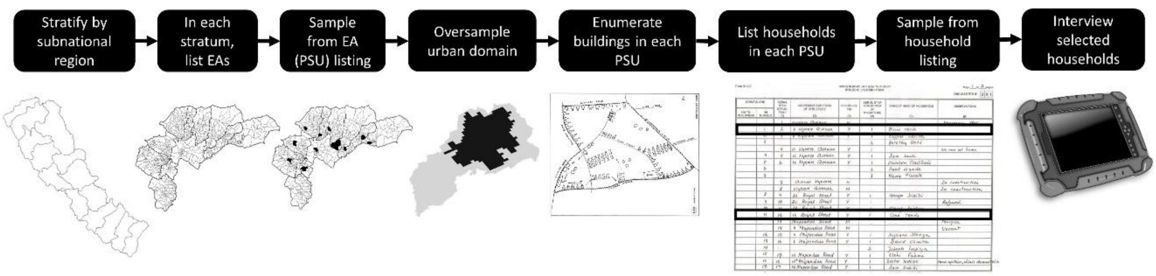


Figure 2. Overview of the standard household survey sampling workflow

Routine national household surveys use the same field implementation methods today as were developed in the 1980s for a, then, majority-rural context. Generally, the buildings in each sampled EA (also called cluster or primary sampling unit – PSU) are manually mapped, and all households are listed by a mapper-lister team several months before an interview team returns and interview 20-30 sampled households per PSU (Figure 2). A team of one mapper and one lister visits each PSU and performs the following. The mapper hand-draws all structures, streets, paths, and landmarks on a blank piece of paper (Grosh and Munoz, 1996; ICF International, 2012a; UNICEF, 2013). Generally, available detailed satellite imagery (e.g., Google Earth) or digital maps (e.g., based on OpenStreetMap) are not used. In routine surveys, if the population identified on the ground is much larger than the target PSU size, the mapper segments the PSU into two or more approximately equal-sized segments in the field, randomly selects one segment using a segmentation form, and then maps the selected segment.

The lister records all dwelling units and households within each mapped structure. According to the DHS and MICS mapping-listing guidelines, all dwellings should be listed in all structures, including mixed residential-commercial and non-residential structures, for example, a guard living inside a factory or in a church (ICF International, 2012a; UNICEF, 2013). Presumably, dwelling and household information is obtained when the lister speaks to residents or their neighbours about the living arrangements in each structure, including questioning guards of commercial and institutional buildings about staff who live on site. However, no explicit guidance or training is provided by the DHS, MICS, or LSMS manuals to operationalize these listing activities (Grosh and Munoz, 1996; ICF International, 2012a; UNICEF, 2013).

The household listing is then used by the survey planning team as the final sample frame to select households to be interviewed. The final step of survey implementation is to send trained interviewers to the field to locate the selected households and administer the survey questionnaire. A great deal of training and guidance is provided to the interview teams to ensure that respondents are well-informed, that confidentiality is achieved during the interview, and that questions are administered and recorded in a standardized way (ICF International, 2012b; UNICEF, 2017). In recent years, the interview process was modernized; interviewers now administer lengthy questionnaires on tablets rather than paper forms in most surveys. After interviews, a number of important steps take place by the survey organizing team to clean, analyse, and distribute the data.

2.4 The unintentionally excluded and masked poor

There are at least five places in this standard household survey workflow where people are unintentionally excluded from the listing (Figure 3). Each of these points is described below and revisited in later chapters. Additionally, there are at least two ways in which urban people, particularly the urban poorest, are misclassified and thus masked in current survey designs (Figure 3). I group these issues by their place in the survey workflow starting with sample *frame* (Chapter 4), then sample *design* (Chapter 5), and finally sample *implementation* (Chapter 6).

The exclusion of homeless, nomadic, and institutional populations from surveys is intentional. This is partly because censuses have well-documented challenges to accurately count these populations (UNSD, 2008, 2017). In addition, survey practitioners often consider homeless, nomadic, and institutional populations as having special needs beyond those of the general residential population, and thus separate surveys with different indicators employing different methods and tools are used, including capture-recapture (Wright and Devine, 1992; Gurgel *et al.*, 2004; Stark *et al.*, 2017), sampling from aggregation points (Peressini, Mcdonald and David, 2010; Troisi *et al.*, 2015), sampling from institutional registrars, and snowball sampling (Mckenzie and Mistiaen, 2009).

However, unintentional exclusion of members of the target population threatens the accuracy of survey results (Biemer, 2010). Furthermore, misclassification of populations and/or failing to disaggregate population figures into meaningful sub-groups masks key survey findings and renders the data unfit for use by decision-makers (Groves and Lyberg, 2010). The Total Survey Error Framework outlines multiple, compounding sources of error in surveys, including errors in sample frame coverage, respondent non-response, sampling, indicator measurement, and data processing (Biemer, 2010). A broader Total Survey Quality Framework encompasses all issues of survey error, but also considers issues of data relevance, usability, accessibility, timeliness, completeness, credibility and comparability (Groves and Lyberg, 2010).

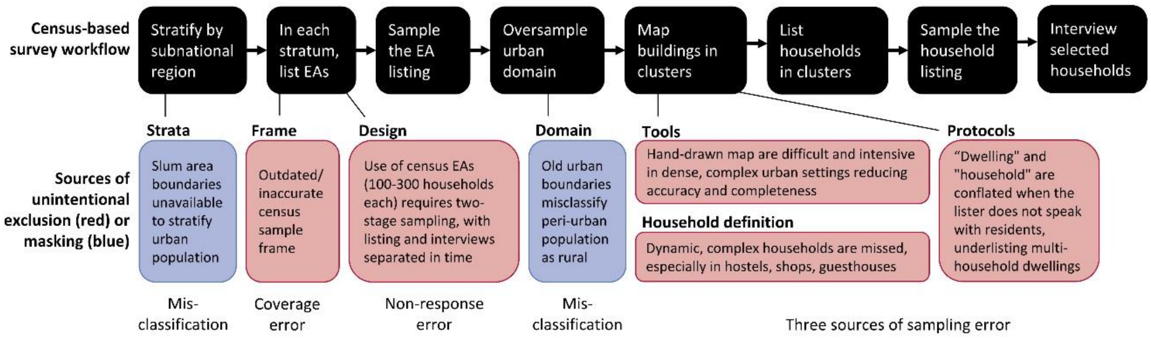


Figure 3. Five ways that poor people may be excluded, and two ways they may be masked, in standard household surveys resulting in bias and/or misclassification
Adapted with permission from (Thomson, Bhattarai, *et al.*, 2020)

2.4.1 Unintentional sample frame exclusion

Exclusion 1: Outdated or inaccurate census sample frame. In outdated or inaccurate sample frames, newly settled communities especially in and around cities, are missing, or communities are over/under-represented due to population shifts driven by births, deaths, and migration resulting in sample frame coverage error. This problem is widespread in LMICs. Weak government, lack of funding, technological limitations, and instability mean that censuses have not been conducted in dozens of LMICs for more than ten years, for example, the D.R. Congo, Somalia, Afghanistan, and Yemen (UNSD, 2019).

On the other end of the spectrum, most LMICs that are stable and developing rapidly have had a recent census; however these datasets become quickly out-of-date due to high levels of migration and urbanization (UNSD, 2005a). For example, the urban population grew by 10 percentage points or more between 2005 and 2015 in Namibia (36.6% to 46.7%), Laos (27.4% to 38.6%), and Thailand (37.5% to 50.4%) (UN-Habitat, 2016). While globalisation drives migration among all socio-economic groups, poor households likely go missing from census sample frames at greater rates than other households in countries where policies lead to mass eviction and displacement of the poor (Carr-Hill, 2013).

In other LMICs, quality of census data is not trusted due to corruption or poor oversight. All of the modern censuses in Nigeria, including the 1962-3, 1973, 1991, and 2006 censuses, have been contentious with accusations of under-counts of rural populations and women, and over-counts in the north of the country (Ahonsi, 1988; Okolo, 1999; Yin, 2007). In these contexts, it is often the poor and marginalized who lack socio-political power who are not accurately recognized and counted (Cobham, 2014), posing problems of under-coverage in the sample frame (Biemer, 2010).

2.4.2 Unintentional sample design exclusion

Exclusion 2: Census EA sample frames require two- or three-stage sampling. Nearly every major household survey conducted since the early 1980s has used census EAs or an equivalent government administrative area dataset as the first-stage sample frame (Table 1). By using census EAs for the initial sample frame, survey designers are forced to use multi-stage cluster sampling techniques because each sample frame unit has a relatively large population. Multi-stage cluster sampling introduces a time lag of several months between the mapping-listing and interview phases (Lê and Verma, 1997), and in practice the lag time can be as long as two years, particularly if the national statistical agency maintains a predefined national sample frame of households for use in multiple surveys (UNSD, 2005a).

Standard household surveys seek to interview 20 to 30 households in each PSU (UNSD, 2005a; ICF International, 2012a). In practice, the number of households per PSU ranges from five (MoH, BoS and ICF Macro, 2010) to 45 (NPC and ICF International, 2014). Typical EAs, on the other hand, are comprised of 100 to 300 households each, and it would not be statistically beneficial or logistically feasible to interview them all (UNSD, 2005a; ICF International, 2012a; UNICEF, 2013). To choose a small sample of households in each PSU, all households must be thoroughly mapped and listed, which is a time-consuming, expensive, and – as I will outline in the following paragraphs – an under-interrogated process. Seasonal or temporary migrants may be systematically under-listed when the migrant household (or neighbour) is not present during the household listing resulting in non-response error. In cases where seasonal and temporary migrant households are mapped and listed, non-response error can be higher in this sub-group if interviews take place during a different season.

While sampling experts expect that the specific households in EAs move or change between decennial censuses, they expect that the total number and type of households in each EA to remain relatively stable over a multi-year period (UNICEF, 2013). Stability of EA population totals might have been a reasonable assumption in 1985, but it is unlikely to hold true in many LMICs undergoing urban transitions today.

2.4.3 Unintentional sample implementation exclusion

Exclusion 3: In practice, households are excluded from surveys resulting in sampling error when non-permanent structures are excluded (Table 2). Manuals by DHS, MICS, and LSMS assume that structures are distinct and have a single use; the manuals provide no specific guidance to map and list non-permanent tent or shack structures in neighbourhoods with permanent buildings, or non-permanent structures occupied by seasonal migrants or construction workers (ICF International, 2012a). If the structures of low-wage workers, including migrants, are not mapped, then their households will be excluded from the household listing.

Exclusion 4: In practice, households are routinely not listed in hostels, shops, and guesthouses resulting in further sampling error. Definitions of a household differ widely by survey and reflect characteristics such as head of household type, sleeping arrangements, shared housekeeping, and/or shared food (Randall *et al.*, 2015) (Table 2). A dwelling is a single living space in a structure which is shared by one or more households (Table 2). Long-term occupants of hostels, shops, and guesthouses are likely excluded from many surveys because their dwellings are not perceived by the mapping-listing team to be residential and thus not approached. Instead, these households might be perceived as settled elsewhere, or otherwise external to the target residential population (UNSD, 2005b).

However, given the high cost of housing in modern cities, a growing number of the working poor take-up permanent residence in non-traditional and shared spaces. People who live in hostels and guest-houses for months at a time are no more homeless or visitors than a young professional who rents a converted garage or shipping container in San Francisco, or who lives on a canal boat in London (Miles, 2016; Parry, 2018). Although these households may move often, they participate in local economies and use local public services and should be considered as part of the general residential population for the purpose of household surveys.

Exclusion 5: In practice, dwellings and households are conflated by the mapper-lister or residents. There are no standard protocols for listers to interact with residents to identify the correct number of dwellings and households per structure (Elsey *et al.*, 2016). The DHS provides the most guidance about mapping-listing, yet still does not provide specific guidance to operationalize definitions and protocols in practice. The LSMS provides the least guidance and only requires a dwelling listing rather than a household listing (Grosh and Munoz, 1996). I summarize the content, and the lack thereof, of DHS, MICS, and LSMS mapping-listing manuals in Table 3.

Questions should be asked of the mapping-listing protocol. In practice, do mappers-listers assume that commercial and institutional buildings are unoccupied? If so, this would lead to an under-listing of cleaning staff, guards, and professionals such as teachers or nurses who live part-time or full-time at their place of work. Also, in practice, do mappers-listers assume the number of dwellings per structure (e.g., one dwelling per structure)? If so, this would result in an under-listing of households in poorer dwellings in many rapidly urbanizing contexts. This is because

Table 2. Definitions of "dwelling" and "household" used in DHS, MICS, and LSMS

Survey	**Structure**	**Dwelling**	**Household**
DHS (ICF International, 2012a)	A free-standing building or other construction that can have one or more dwelling units for residential or commercial use. Residential structures can have one or more dwelling units (for example: single house, apartment structure).	A room or a group of rooms normally intended as a residence for one household (for example: a single house, an apartment, a group of rooms in a house); a dwelling unit can also have more than one household.	A person or a group of related or unrelated persons, who live together in the same dwelling unit, who acknowledge one adult male or female 15 years old or older as the head of the household, who share the same housekeeping arrangements, and are considered as one unit. In some cases, one may find a group of people living together in the same house, but each person has separate eating arrangements; they should be counted as separate one-person households. Collective living arrangements such as army camps, boarding schools, or prisons will not be considered as households. Examples of households are: • a man with his wife or his wives with or without children • a man with his wife or his wives, his children and his parents • a man with his wife or his wives, his married children living together for some social or economic reasons (the group recognize one person as household head) • a widowed or divorced man or woman with or without children
MICS (UNICEF, 2013)	A free-standing building that can have one or more dwellings for residential or commercial use. Residential structures can have one or more dwelling units (e.g., a single house or an apartment building).	A room or a group of rooms normally intended as a place of residence for one household (e.g., a single house, an apartment, or a group of rooms in a house). However, a dwelling unit can also be shared by more than one household.	A person or a group of related or unrelated persons, who live together in the same dwelling unit, who share common living arrangements, who acknowledge the same person as the household head, who eat together and are considered as one unit.
LSMS (Grosh and Munoz, 1996)	Not clear. “Dwelling” and “building” often used interchangeably.	A group of rooms or a single room occupied or intended for occupancy as separate living quarters by a family or some other group of persons living together, or by a person living alone.	A group of people who share a roof and a cooking pot.

developers add new levels to the tops of structures and landlords carve dwellings into smaller units to meet demands for affordable housing. These scenarios were encountered routinely during a household survey in Kathmandu, which will be discussed in Chapter 6.

As a result of limited guidance, mappers-listers are very unlikely to have standardized, detailed interactions with residents. Instead, mappers-listers may make assumptions about the types of structures that are occupied, number of dwellings per structure, and the number of households per dwelling. When mapper-listers hold brief conversations with residents, residents may not fully understand the distinction between households and dwellings, and thus a multi-household dwelling may be listed as one household. All of these issues would result in sampling error.

Multi-household dwellings are increasingly common in LMIC cities among the poor. While the LSMS manual reports, “the average number of households per dwelling ranges from 0.9 to 1.1 in most countries” (Grosh and Munoz, 1996), this text is more than 20 years old, and misleading if it averages across rural and urban settings. The average number of households per dwelling today is likely to vary widely between urban and rural settings, and by country. A 2015 urban health survey in Kathmandu, Nepal, for example, reported a median of 2 households per dwelling, with a range of 1 to 4 household per dwelling across 72 randomly selected PSUs (Elsey *et al.*, 2016).

The lack of materials to train mappers and listers stands in stark contrast to the training of interviewers. The Training Field Staff for DHS Surveys manual includes a template interviewer job description, a candidate assessment screening tool, a template training schedule, content and exercises for the training, sample tests, and field evaluation criteria (ICF Macro, 2009). Additionally, template training manuals for the interviewers are provided by DHS and MICS covering methods to build rapport with respondents, with detailed procedures to locate a dwelling, approach a household, identify an eligible respondent, handle refusals, and complete each question of the questionnaire on a tablet versus back-up paper form (ICF International, 2012b; UNICEF, 2017). Table 3 includes similar materials that mappers-listers and their trainers would need to operationalize their work, but which are largely absent in routine survey programmes.

Table 3. Summary of DHS, MICS, and LSMS protocols for the survey mapping-listing process

Three green blocks indicate that the manual has an operational protocol, two blocks indicates a partially operational protocol, and one block indicates a summary description of a protocol that is not operational.

Criteria	**DHS**	**MICS**	**LSMS**
GENERAL			
Overview of survey aims and design	■■■	■■■	■■■
Overview of the mapping-listing process	■■■	■■■	■□□
Definitions of key terms	■■■	■■■	■□□
Two-person mapping-listing team specified	■■■	■■■	□□□
Recommend # of mapping-listing teams	□□□	□□□	□□□
General responsibilities given, by role	■■■	■■■	□□□
TRAINING MATERIALS FOR TRAINERS			
Candidate profile	□□□	□□□	□□□
Candidate assessment tools	□□□	□□□	□□□
Template training schedule	□□□	□□□	□□□
Template training content	□□□	□□□	□□□
Template location (of cluster) form	■■■	□□□	□□□
Template structure mapping form	■■■	■■■	□□□
Template dwelling/household listing form	■■■	■■■	■■■
Template segmentation form	■■■	■■■	□□□
Materials checklist for fieldwork	□□□	■■■	□□□
Training evaluation criteria / tool	□□□	□□□	□□□
TRAINING MATERIALS FOR MAPPERS-LISTERS			
How to build rapport with local authorities	□□□	□□□	□□□
How to build rapport with residents	□□□	□□□	□□□
Protocol to prepare location map	■□□	■□□	■□□
Example (cluster) location map	■■■	■■■	□□□
Protocol to locate a cluster	□□□	□□□	□□□
Protocol to record cluster GPS coordinate	■■■	□□□	■□□
Protocol to segment large cluster	■■□	■□□	■□□
Example completed cluster segmentation	■■■	□□□	□□□
Protocol to locate structure	■□□	■□□	□□□
Protocol to record structure	■■□	■□□	□□□
Example completed structure map	■■■	■■■	□□□
Protocol to locate dwelling	□□□	□□□	□□□
Protocol to record dwelling	■□□	■□□	■□□
Protocol to locate household	□□□	□□□	
Protocol to record household	■■■	■■□	
Example completed listing	■■■	□□□	□□□
Protocol to split mapping & listing activities	□□□	□□□	□□□
Protocol for supervisor quality control	■□□	□□□	□□□
GUIDANCE FOR PLANNING TEAM			
Protocol to securely store listing	□□□	□□□	□□□
Protocol to select household/dwelling sample	■■■	■■□	■■■

2.4.4 Masking the urban poor by design

Masking the poorest 1: Slum area boundaries are not available to stratify urban populations. Slums are a growing reality in LMIC cities, and are associated with disparities in individual health due to lack of basic water and sanitation systems, security of tenure, and/or access to government services (Ezeh *et al.*, 2017). While surveys and censuses have long reported disaggregated statistics for urban and rural populations due to major differences in their health and wellbeing, such disaggregation is extremely rare for urban slum and urban non-slum populations, masking stark disparities between these groups. The 2005-06 and 2015-16 India DHSs are among the few surveys to stratify urban slum and non-slum areas and report disaggregated results (IIPS and Macro International, 2007; IIPS and ICF International, 2017). A number of studies in LMICs have found that urban slum dwellers have similar, or even poorer, child health outcomes than the poorest rural populations (Agarwal, 2011; Ezeh *et al.*, 2017). Given the importance of slum/non-slum disaggregated statistics for monitoring SDGs and making city-level investment and development decisions (Lilford *et al.*, 2017; UN-DESA, 2018; Thomson, Linard, *et al.*, 2019), the Total Survey Quality framework indicates that absence of slum/non-slum disaggregation in survey statistics threatens the relevance of survey results (Biemer, 2010; Groves and Lyberg, 2010). The main reason that urban estimates are not disaggregated, is that national statistical offices lack a common slum area definition, and the capacity to map all slum areas in all urban areas in LMICs is limited (Lilford *et al.*, 2017).

Computer generated boundaries from high resolution satellite imagery around areas of concentrated, disorganized, small buildings has yielded several slum area maps in LMICs; however, these maps are limited in their coverage by the cost of high-resolution satellite imagery and need for extensive computing power (Kuffer, Pfeffer and Sliuzas, 2016). Satellite imagery-based slum area maps often lack field validation, and only reflect one characteristic of a slum: unplanned, densely arranged structures. In select countries such as Bangladesh (Angeles *et al.*, 2009) and Brazil (Snyder *et al.*, 2014), the government has been involved with lengthy and costly field-based slum mapping efforts. Given the growing recognition that urban slum populations comprise an increasing and substantial portion of urban populations in LMICs, the international community is beginning to push for scalable efforts to map slum area boundaries to disaggregate urban data in future censuses and surveys (Lilford *et al.*, 2017).

Masking the poorest 2: Outdated urban boundaries misclassify peri-urban populations as rural. It is common for the urban poorest to reside in peri-urban areas in order to access the city's economy and services whilst maintaining affordable housing. Around rapidly expanding cities, basic water and sanitation infrastructure and government facilities are not installed at the same pace as new housing, resulting in concentrated slums (UN-Habitat, 2016). Depending on the last time that official urban and rural area boundaries were updated, households located in these peri-urban regions may function as part of the city but be misclassified as rural during a survey threatening relevance of survey results for decision-making (Groves and Lyberg, 2010). Counting urban slum dwellers as rural masks their unique urban poverty which results from living in high density conditions and being more dependent on a cash economy than rural households.

2.5 Gridded population sampling

This section provides a brief explanation of gridded population data sources and sampling approaches for context, though gridded population data and gridded population sampling are detailed in Chapters 4 and 5, respectively. Gridded population sampling refers to any survey sample selected from a gridded population dataset. Gridded population sampling is an emerging field that began around 2010 after gridded population datasets had been freely and publically available for LMICs for several years.

2.5.1 Gridded population data

Gridded population data are estimates of the total population in small grid cells derived from a geo-statistical model using a number of spatial datasets (see Figure 4). The grid cells generally range in size from 100 metres by 100 metres to 1 kilometre by 1 kilometre, and many of these disaggregated population estimates are free and publicly available.

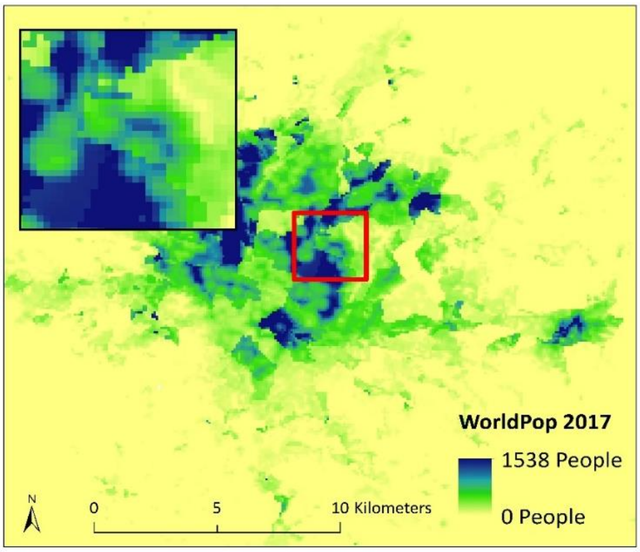


Figure 4. Example 100 metre by 100 metre gridded population dataset, Kathmandu, Nepal

Data from WorldPop (WorldPop, 2019)

Gridded population data provide analysts the flexibility to model population phenomena at a local level across a surface (Alegana *et al.*, 2016; Perkins *et al.*, 2016), and to re-aggregate population estimates to new geographic units that are more relevant to the study (Tatem *et al.*, 2014; Tatem, 2017). For example, in countries where population micro data (individual- or household-level data) are available, including most European countries, aggregation of point-level population data to grid cells maintains a high level of detail whilst preserving confidentiality of individuals (European Commission, 2012). Decoupling population counts from administrative boundaries – which may vary in shape and size, and may not follow environmental, cultural, or political features – means that population counts can used in a wide range of health (Perkins *et al.*, 2016; Tatem *et al.*, 2016), environmental (Early *et al.*, 2016), and economic (Buhaug and Urdal, 2013) applications.

Data sources used to model gridded populations vary and include: total population counts (e.g. census data) or population counts in select small areas (e.g. microcensuses or survey household listings); geographic vector data such as roads, water bodies, and building locations; satellite raster data such as temperature, elevation, and night-time lights; and other modelled spatial data layers such as urban extent boundaries. A number of gridded population datasets now incorporate "big data" such as geo-tweets or Facebook account locations (Facebook Connectivity Lab and CIESIN, 2016; Patel *et al.*, 2017) and aggregated mobile phone call detail records (Deville *et al.*, 2014; Lu *et al.*, 2016; Wilson *et al.*, 2016) as ancillary data.

Geo-statistical techniques used to model gridded population data range in complexity from uniform disaggregation of total population counts (UNEP, 2006; Doxsey-Whitfield *et al.*, 2015); to informed-disaggregation of total population counts using areal weighting with ancillary data (Balk *et al.*, 2005; Facebook Connectivity Lab and CIESIN, 2016; Pesaresi, Ehrlich, *et al.*, 2016); to advanced disaggregation using complex models (Dobson *et al.*, 2000; Azar *et al.*, 2010, 2013; Stevens *et al.*, 2015). All of the aforementioned modelling techniques are considered "top-down" approaches, as they involve disaggregating total population counts to small grid cells.

The benefit of top-down gridded population models is the possibility of improving the relative distribution and/or detail of population dataset within smaller areas, though this is constrained to circumstances when population totals are reasonably accurate. Thus, "top-down" gridded population models derived from 15-year-old census data may result in an improved population distribution, but population totals would be incorrect. This hypothesis will be evaluated in Chapter 4. There may be ways to improve the accuracy of these gridded population distributions before sampling, for example, by aggregating to 500 metre by 500 metre grid cells, also addressed in Chapter 4.

Additional “bottom-up” modelling techniques are being developed to model population totals from population counts in small areas, and are thus attractive when census data are not available. Bottom-up gridded population models use the relationship between population density in a selection of small areas (e.g., microcensus) and a number of spatial covariates to predict population counts in unmeasured areas of the country (Weber *et al.*, 2018).

2.5.2 New inclusive survey methods

New survey methods and tools might improve the accuracy of household surveys (Figure 5). When census data are outdated or inaccurate, gridded population datasets might instead be used as a sample frame (Alkire and Samman, 2014). Even “top-down” gridded population datasets derived from outdated or inaccurate census data, might result in more accurate samples of PSUs if the relative distribution of population density is more detailed and accurate compared to the census data.

Area-microcensus sampling (also called one-stage sampling or modified cluster design) means that all households in a small area are sampled. Gridded population sample frames with 100 metre by 100 metre grid cells enable area-microcensus sampling, and eliminate the time gap between the household listing and interviewers, potentially improving representation of mobile populations. Interviewing PSUs across different seasons could additionally improve response rates of seasonal migrants. Chapter 6 compares the number and types of households that were sampled in an area-microcensus versus two-stage survey conducted concurrently in Kathmandu, Nepal in 2017.

Finally, robust mapper-lister protocols to interact with residents in a standardized, detailed way are needed in household surveys of any design. These protocols should include clear definitions of structures, dwellings, and households which are present in complex urban settings, and guidance about how to approach residents of all structures to ensure that informal and atypical households are not omitted from the listing.

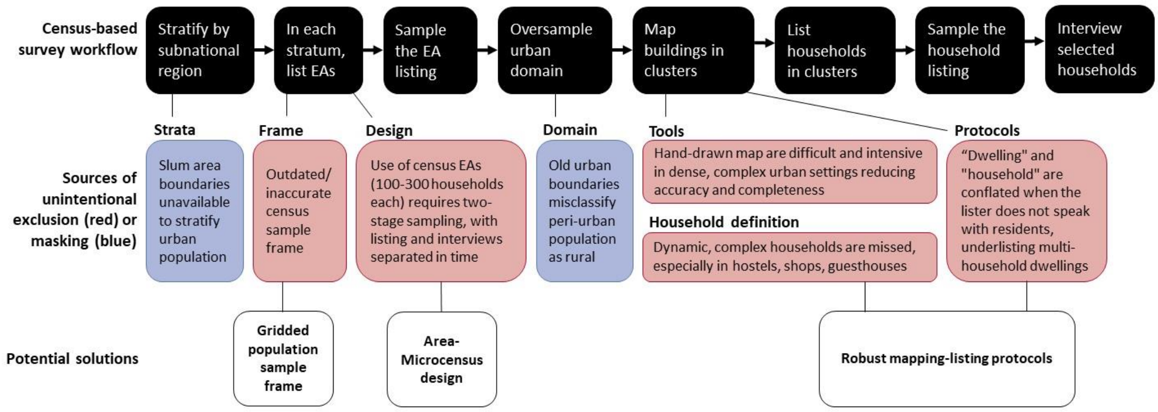


Figure 5. Potential tools and methods to improve accuracy of household surveys

Adapted with permission from (Dana R. Thomson *et al.*, 2020)

2.5.3 Gridded population surveys

Selecting survey samples from "top-down" or "bottom-up" gridded population data is worth considering in LMICs with outdated or inaccurate census data. More than two dozen national gridded population surveys have been conducted across diverse settings including Iraq (Galway *et al.*, 2012; Hagopian *et al.*, 2013), Somalia (Pape and Wollburg, 2019), Colombia (Cajka *et al.*, 2018), and Thailand (Cajka *et al.*, 2018). Multiple sub-national surveys have also been conducted in settings that include eastern DR Congo (Thomson *et al.*, 2012), Kathmandu Valley Nepal (Elsey *et al.*, 2016), and Chin State Myanmar (Sollom *et al.*, 2011). A review of gridded population surveys is provided in Chapter 5.

All of the aforementioned gridded population surveys were motivated for one of two reasons. The first reason, and by far the most common reason, was that the census sample frame was unavailable, outdated, or had known inaccuracies. The second reason was that the survey was based in a complex urban environment where an area-microcensus sampling design was preferred to ensure representation of vulnerable and mobile populations.

2.6 Change in survey accuracy as a result of population trends

It is likely that household survey samples were more accurate at the time that DHS, MICS, and LSMS were established compared to today. It is not that standard survey methods are inherently biased, but rather than population characteristics and behaviour have changed substantially over the last four decades resulting in an increasing proportion of people who fall through the cracks of dated survey methods and tools.

Most LMICs are in the midst of urban transitions, or will be soon (UN-DESA, 2014). Urban transitions are comprised of both a demographic transition – in terms of increased population density, change in household structure and

size, and change of population composition resulting from reduced mortality and fertility – as well as a mobility transition (Woud, Jan and Hayami, 1990). Urbanization has been associated with some of the greatest achievements in human history, including reductions in mortality, extended life spans, production of material wealth and wellbeing, and harnessing of diverse minds to create social and economic movements (UN-DESA, 2014; UN-Habitat, 2016). However, urbanization is also closely linked with socioeconomic inequalities that trap generations of families in cycles of crushing poverty and insecurity (Diaz *et al.*, 2017). Socioeconomic inequalities within countries are more pronounced today than at any other time in the last 30 years, particularly in LMICs (UN-Habitat, 2016).

Future population growth is expected to be concentrated in LMICs. The global population doubled between 1970 and 2015 from 3.7 to 7.4 billion, and it is projected to reach 9.8 billion by 2050 (UN-DESA, 2017). Asia has been, and will remain, the most populous region (Figure 6), with five countries - Bangladesh, China, India, Indonesia, and Pakistan - accounting for 45% of the global population (Roser, Ritchie and Ortiz-Ospina, 2016). However, more than half of the projected global population growth between now and 2050 is expected to occur in Africa where 33 the world's 47 least developed countries (LDC) are located (UN-DESA, 2017). The United Nations (UN) designates LDCs as low-income countries which face major structural obstacles to sustainable development.

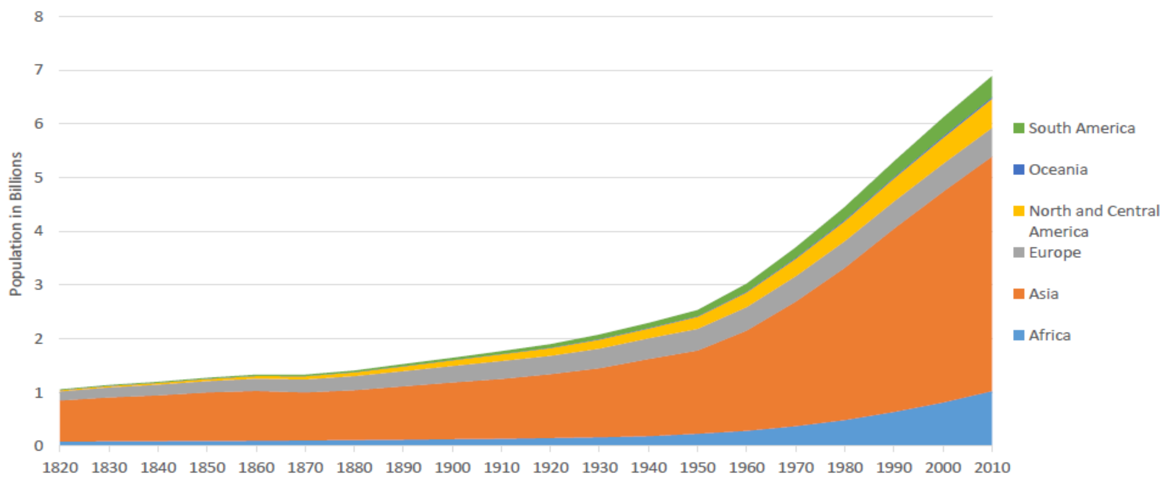


Figure 6. World population by world region, 1820-2010

Data from Our World in Data (Roser, Ritchie and Ortiz-Ospina, 2016)

Future population growth in LMICs will be concentrated in urban areas. In 2007, the world's population became majority urban with most of the urban population living in Asia, followed by Europe, and the Americas (UN-DESA, 2014). Despite high rates of urbanization, most LMICs are still majority rural today, but this is changing rapidly as LMICs enter urban transitions. By 2030, a majority of the population in LMICs will live in urban rather than rural areas, and by 2050, two-thirds of the population in LMICs is expected to be urban (UN-DESA, 2014). An estimated 2.5 billion people will be added to the planet by 2050, with 90% of that population increase concentrated in Asian and African cities alone (UN-DESA, 2014).

While *rates* of urban growth in LMICs are currently on par with, or lower than, urban growth rates previously observed in high-income countries (Satterthwaite, 2010), the *absolute number* of people being added to urban areas today creates entirely new scenarios of urbanization. In 2014, the world had 28 cities with 10 million or more residents each, and by 2030, the number of these so called "megacities" is expected to double with most new megacities being added in LMICs (UN-DESA, 2014). While some of this rapid urban growth can be explained by changes in the definitions of city boundaries and the use of old projected population data (Satterthwaite, 2010), undeniably, the absolute number of people moving to urban centres in LMICs over the last two decades is unprecedented. Since the early 1990s, a majority of megacities in Asia were thought to have added more than 400,000 people per year on average (Figure 7) (UN-DESA, 2019). The UN projects that Lagos (Nigeria), Delhi (India), and Dhaka (Bangladesh) will each add more than 700,000 people per year on average through 2030 (Figure 7) (UN-DESA, 2019).

Megacities, however, are not the fastest growing cities, nor do they represent a majority of the world's urban population. Most of the urban growth over the next 15 years will be in medium and small cities of less than 1 million people in Africa and Asia (UN-DESA, 2014). Even if population projections are high, measurements of urban extents and informal housing settlements via satellite imagery have recorded rapid sprawl of city boundaries and concentrations of informal housing in LMIC cities (Pesaresi, Melchiorri, *et al.*, 2016), both of which place excessive demands on city officials to extend city infrastructure and services (UN-Habitat, 2016). It is in this context of rapid urbanisation that updated survey methods and tools are needed to accurately measure modern LMIC populations.

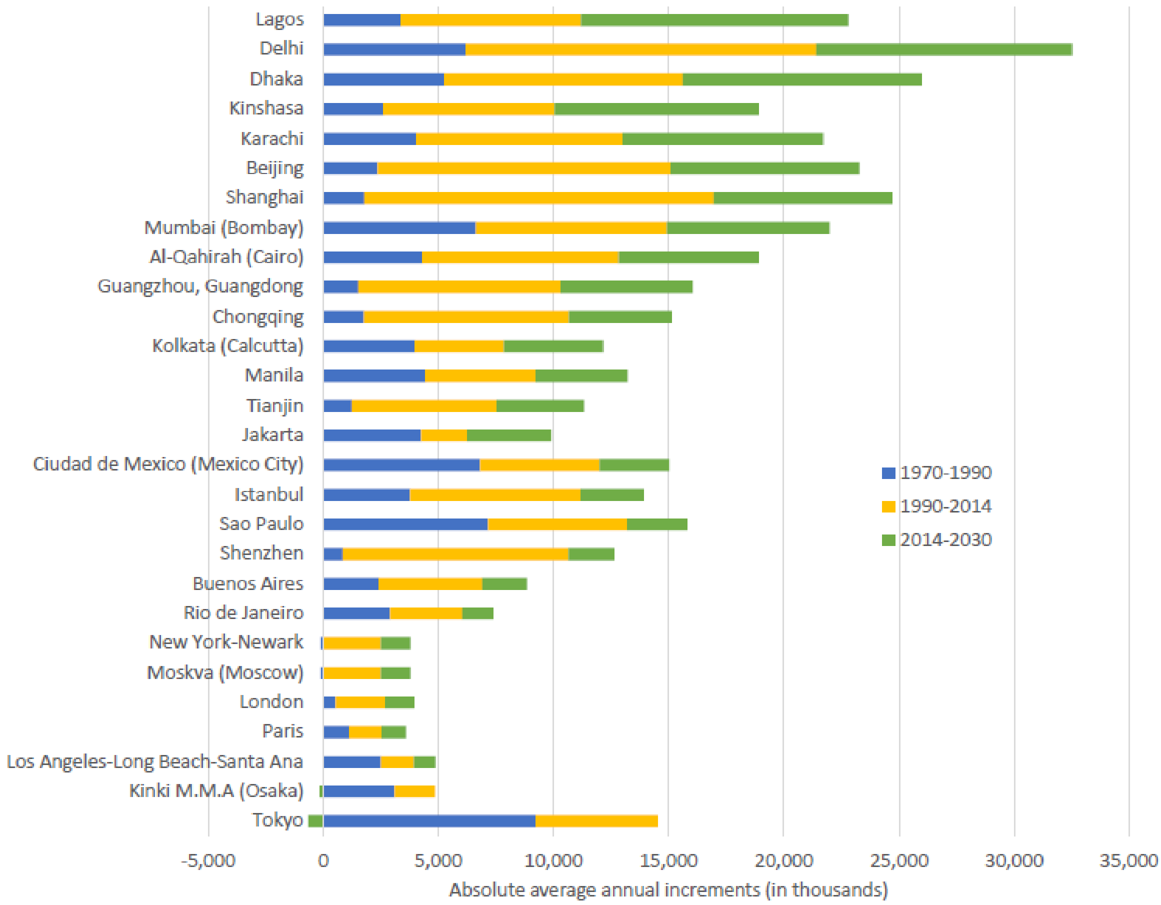


Figure 7. Absolute average population increase in 28 megacities for periods 1970-1990, 1990-2014 and 2014-2030

Data from World Urbanization Prospects: 2014 Revision (UN-DESA, 2014)

2.7 Aims and Objectives

This thesis aims to evaluate the accuracy and feasibility of gridded population sampling when census sample frames are outdated or inaccurate.

Chapter 3: Simulation. Objectives:

- Simulate a geo-located synthetic "true" population with demographic and health characteristics for a LMIC
- From the synthetic population, simulate several censuses with realistic patterns of outdatedness and inaccuracy due to under-counts

Chapter 4: Frames. Objectives:

- For each simulated census, generate a top-down gridded population dataset

Chapter 2

- Evaluate cell-level accuracy of gridded population datasets derived from outdated-inaccurate census data at multiple geographic scales by comparing to the "true" synthetic population

Chapter 5: Sampling. Objective:

- Evaluate the accuracy of survey results that are sampled from equivalent outdated-inaccurate censuses and gridded population datasets, using characteristics in the "true" synthetic population for comparison

Chapter 6: Implementation. Objectives:

- Describe the implementation of a real-world gridded population survey in a LMIC
- Assess the feasibility of the implementation methods based on qualitative data from the survey planning team and field teams
- Compare the types of households listed and interviewed in an area-microcensus versus two-stage sample design

Chapter 3: Simulation

3.1 Overview

There are two ways in which census sample frame coverage errors directly bias surveys. First, by being grossly outdated, and second by being poorly conducted, both of which result in inaccuracies that overwhelmingly omit the poorest populations in low- and middle-income countries (LMICs) (Cobham, 2014). Census inaccuracies can be deliberate or unintentional. Deliberate inaccuracies may be introduced to gain or maintain socio-political power; for example, manipulating population totals to influence where district boundaries are drawn, the number of seats in parliament, or national budget allocation (Yin, 2007). Likewise, census inaccuracies may be introduced to prevent the allocation of resources to new political parties or groups. For example, the refusal to recognize slum areas prevents governments from being obliged to spend limited public funds on infrastructure and services in communities who generate limited tax revenue, and ensures that those communities do not gain the political power to demand otherwise (Agarwal and Taneja, 2005; Subbaraman *et al.*, 2012).

Recent examples of deliberate census manipulation come from Nigeria, India, and Ethiopia where censuses were criticized for inflating population counts in sub-regions, widespread omission of “illegal” slum residents, or wilful negligence during data collection (Yin, 2007; Carr-Hill, 2013; Bekele, 2017). Falsification of population counts has long been a problem in Nigerian censuses (Ahonsi, 1988; Okolo, 1999), and the 2006 census was no exception with accusations that numbers in the north of the country were vastly inflated for political and economic gains (Yin, 2007). A comparison of the 2006 Nigeria census with a parallel census of Lagos city conducted by the city government revealed an under-count of 8.5 million inhabitants in the national census, though political motives may have been involved in the city census as well (Yin, 2007). In Ethiopia, independent analyses of the 1994 and 2007 censuses identified major anomalies in reported age-sex demographics, suggesting substantial problems with the quality of data collection (Randall and Coast, 2016; Bekele, 2017). In India, the 2001 census was widely criticized for excluding unrecognized informal settlements and people living in poor-quality inner city housing who account for up to half of poor urban households (Agarwal, 2011; Carr-Hill, 2013).

While deliberate inaccuracies may make headlines, unintentional census under-counts are perhaps more common in LMICs, especially among informally-housed and mobile populations due to the complexity of classifying their dwellings and households (Lavely, 2001). A well-conducted census is expected to have under-counts up to 5% due to logistical difficulties, and should be adjusted based on post-enumeration surveys (PES) in randomly selected enumeration areas (EAs), especially if the under-counts are concentrated in certain sub-groups (UNSD, 2017). Unfortunately, relatively few LMICs conduct PESs (UNSD, 2010).

Censuses generally count institutional, homeless, and nomadic populations separately from the main census. Institutional populations are enumerated from institutional registrars, and special enumeration activities are conducted to count homeless and nomadic populations because they move often, and generally do not live in permanent structures. In some countries, certain refugees or internal displaced people (IDPs) may be omitted by design as well (Cobham, 2014). However, many censuses in LMICs are not resourced to carry out additional enumeration activities, so homeless and nomadic populations either go uncounted or are estimated with crude methods (UNSD, 2010). Due to difficulties in counting institutional, homeless and nomadic populations in censuses, large-scale surveys in LMICs are designed to only represent permanent and semi-permanent conventional households. Technically, survey results are not biased if populations missing from the sample frame are excluded by design; however, these exclusions have important policy implications.

I chose to simulate a “true” population rather than use an existing census for two reasons. The primary reason is that no LMIC census publicly releases the type of data that I need. I need the geographic point location for each household, as well as the 100% census for a region to be able to measure the reference (true) population for accuracy assessments in Chapters 4 and 5. Furthermore, censuses, particularly in LMICs, inevitably exclude some vulnerable and mobile populations, and few LMIC censuses have performed a PES to appropriately adjust population counts.

In this chapter, I simulate a realistic “true” 2016 population in Khomas, the capital city region of Namibia, and spatially allocate simulated households to actual 2016 building locations according to satellite imagery. From these “true” households, I simulate 16 realistic census datasets with varying degrees of outdatedness and inaccuracies. Outdatedness is simulated by removing households assigned to buildings that were not present in satellite imagery in 2011, 2006, and 2001. Census inaccuracy is simulated by removing households based on rates of census under-counts identified in a literature review. In the following chapters I use the simulated census datasets and the “true” reference population to assess cell-level accuracy of gridded population datasets derived from censuses, and the accuracy of subsequent gridded population survey samples.

3.1.1 Aims

Aim 1: Simulate realistic household locations and characteristics based on the geography of Khomas, Namibia, utilising 2011 Namibia census and 2013 Demographic and Health Survey (DHS) data.

Aim 2: Conduct a literature review to determine rates of under-counting in LMIC censuses among different sub-populations.

Aim 3: Simulate several census datasets from the realistic population by applying rates of outdatedness and inaccuracy, and aggregating to enumeration (EA) boundaries.

3.2 Requirements

To achieve the research aims of this PhD, I outline four requirements for the simulated realistic population.

First, the simulated "real" population should be geo-located to an actual place where covariate datasets exist with all of their own imperfections to be able to generate gridded population datasets in later analyses. While I could have chosen to simulate a population in a theoretical place, and simulate dozens of spatial covariates typically used in gridded population models, I want to ensure that the covariate datasets are not unrealistically perfect. Like census data, covariate datasets such as road networks and building footprints are subject to inaccuracies and outdatedness, and these imperfections need to be reflected in the analysis if the results of this thesis are to be relevant to researchers, practitioners, and decision-makers.

The second requirement is that the simulated "real" population needs to include sizable numbers of vulnerable and mobile populations to reflect the different components of missingness and outdatedness described above. This will enable the evaluation of census population missingness and outdatedness on gridded population estimates, and ultimately on gridded population surveys.

A third requirement is that the setting should have highly concentrated populations as well as vast unsettled areas. A key difference among gridded population modelling approaches is that some models estimate population everywhere including in deserts and forests (Doxsey-Whitfield *et al.*, 2015; Stevens *et al.*, 2015), while other models constrain population estimates to areas classified as settled, and are thus subject to inaccuracies in the settlement layer (Dobson *et al.*, 2000; Pesaresi, Ehrlich, *et al.*, 2016). The former approach is likely to estimate population where none exists and underestimate population in settled areas, while the latter approach is likely to estimate zero population in small rural settlements and overestimate counts in urban areas. Thus, to evaluate accuracy of gridded population datasets in settled and unsettled areas, settings that contain both major cities and large deserts or forests are considered.

Lastly, for the simulated population to be realistic, it needs to be based on high quality and very detailed census and survey data, thus countries with a record for high data quality are considered.

3.3 Setting

I selected Namibia for this simulation because the government has produced numerous high-quality population datasets and the geography is defined by vast unsettled areas. I focus specifically on Khomas, one of Namibia's 13 regions, located in the centre-south of the country (Figure 8). Khomas covers an area of 37,007 square kilometres and is home to roughly 342,000 people according to the 2011 census (NSA, 2011b). Simulating a population for the whole country was not feasible, as building locations had to be manually digitized to locate simulated households.

Windhoek, Namibia's capital, and its suburbs are where essentially all of the population in Khomas are located; the surrounding areas are mainly unpopulated desert with a few small hamlets. Namibia, like many other countries that inherited colonial boundaries, placed restrictions on freedom of movement until independence in 1990 (Newaya, 2010). After independence, vast numbers of the population migrated to Windhoek, exaggerating rural-to-urban migration patterns observed globally during this time period (IOM, 2015; Lai *et al.*, 2019). Windhoek is also a destination for immigrants from neighbouring countries including financially unstable Zimbabwe (IOM, 2015; WorldPop, 2016). The population of the Windhoek metropolitan area grew by a staggering 37% between the 2001 and 2011 censuses (NSA, 2011b), with much of that growth in informal settlements (Newaya, 2010).

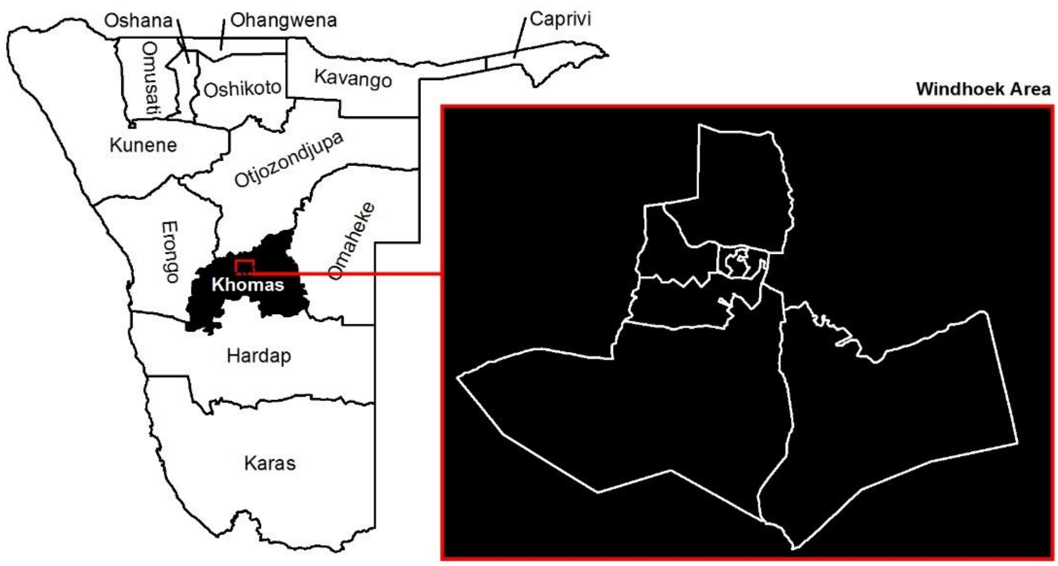


Figure 8. Location of Khomas region in Namibia, and of constituencies in Windhoek area

Although a PES was not conducted, the 2011 Namibia census meets the UN recommendations for a high quality census (UNSD, 2019), including checks and editing during paper-based data collection, checks and editing during electronic data entry, imputation of missing and implausible values for select variables such as person’s age, and collection of homeless and institutional population counts (NSA, 2013). The Namibia National Statistics Agency (NSA) makes its census EA boundaries available (NSA, 2011a), and distributes census microdata samples in the IPUMS database with detailed documentation (Minnesota Population Center, 2018). Furthermore, the NSA publishes routine household survey data including DHSs (MoHSS and ICF International, 2014).

3.4 Simulating a “real” population by household location

This section describes the methods I used to simulate a “real” population of every person and household in Khomas in 2016, and the methods I used to link synthetic households to actual building point locations digitized from 2016 satellite imagery. Results are presented and discussed for each phase of the process.

3.4.1.1 Approaches to population simulation

Several simulation approaches are available to construct a synthetic population and its spatial distribution. The most common approach is to expand or reweight observations from a survey of individuals to meet totals and marginal distributions in more aggregated areal units reported in a census. These approaches include iterative proportional fitting (IPF) (Birkin and Clarke, 1988, 1989), and combinatorial optimization procedures such as simulated annealing (SA) (Ballas *et al.*, 2007) or quota sampling (Farrell, Morrissey and O'Donoghue, 2012). IPF models incrementally improve the fit of a joint probability distribution of person- or household-level attributes (e.g., from a household survey) subject to known joint probabilities of these attributes (e.g., from an aggregated census). Combinatorial optimization procedures follow a similar approach but prevent sub-optimal combinations of attributes in the simulated dataset. These models can be implemented with open-source software such as the *simPop* R package (Templ *et al.*, 2017). Other approaches to population simulation include agent-based models of agents (i.e., individual people) with key attributes and relationships (Macal, 2016; Chapuis *et al.*, 2018), including models that incorporate agent interactions in space (Heppenstall, Malleson and Crooks, 2016). The vast majority of simulation models produce synthetic population counts in small output areas, such as census EAs.

Some have gone further to assign synthetic households to spatial point locations. To create the RTI U.S. Synthetic Household Population, RTI assigns households from US census tracts or block to random point locations alongside roads for the entire country (RTI International, 2016). In the U.S., this can result in a reasonably accurate simulation of households by point location because nearly every household is located along a road, and road datasets are very complete. However, in most LMICs, road datasets are not complete and many households are located far from roads with only foot access. To simulate spatial locations of households globally, the Synthetic Populations and Ecosystems of the World (SPEW) project creates a random point for each household within in the smallest administrative unit available (Richardson *et al.*, 2017). While this simplistic assumption about spatial location is sensible for their global models of infectious disease, it is not sufficiently detailed or realistic to simulate household locations for this analysis.

To evaluate the accuracy of gridded population sample frames in the next chapter (Chapter 4), I model 100 metre by 100 metre gridded population estimates from each of the simulated census datasets created in this chapter. Gridded population models use dozens of fine scale spatial covariates that together identify areas of likely human activity on the Earth's surface. For this reason, it is important that simulated household locations are as realistic as possible so I can calculate a realistic reference population in every 100 metre by 100 meter grid cell to assess gridded population model accuracy. I decided that the best way to approach this was to digitize actual building locations.

3.4.1.2 Approach used

To simulate a realistic population in Khomas, Namibia, I simulated a synthetic population by second-level administrative boundaries (i.e., constituencies), then distributed synthetic households to actual building point locations that I digitized over satellite imagery. The allocation of simulated households to point locations was based on socio-demographic household types and environmental covariates. I worked closely with Dr. Lieke Kools and Dr. Warren C. Jochem to develop this methodology. Our methods are described in detail in the literature, and demonstrated for Oshikoto, Namibia (Thomson, Kools and Jochem, 2018). In this chapter, I repeat the simulation for Khomas in three phases, and provide interim outputs, a summary of the final simulation, and quality assessments of the final simulation. These simulation methods should not be considered a unique contribution of this thesis as they were conceived and developed jointly with Dr. Kools and Dr. Jochem.

3.4.1.3 Data

All input datasets used in the simulation are publicly available. These include the 2011 Namibia census 20% microdata sample (NSA, 2013) and 2011 Namibia census enumeration area boundaries (NSA, 2011a) available by request from the Namibia NSA, as well as the 2013 Namibia DHS recode files and geo-displaced cluster coordinates available by request from the DHS Project (ICF International, 2020). The simulation also used 2014-2016 Quickbird (30cm) satellite imagery available in ArcGIS10.5 (Maxar, 2019); 2004-2016 Maxar and SPOT (40cm) satellite imagery available in Google Earth Pro (Google LLC, 2019), and spatial covariates from the WorldPop-Global project including land cover type, night-time lights intensity, and health facility locations (Lloyd, Sorichetta and Tatem, 2017). All datasets used in this simulation are summarized in Table 4.

I expected environmental covariates to have a relationship with both population density and population characteristics across space, and thus used them to allocate simulated household records to digitized building locations. To prevent overtraining of models, I used different microdata sources to simulate the population (20% microdata census sample) and to train household-to-point allocation models (DHS sample).

Table 4. Data sources for simulated population in Khomas, Namibia

Adopted with permission from (Thomson, Kools and Jochem, 2018)

Short name	Long name	Source, original unit	Output unit
Population			
dhs_hh	Individual recode file summarized by household	2013 Demographic and Health Survey (1)	region
dhs_geo	Geo-displaced cluster coordinates	2013 Demographic and Health Survey (1)	coordinate (cluster)
census_housing, census_person	20% microdata census sample	2011 Namibia Statistics Agency (2)	constituency
census_report	Final census report	2011 Namibia Statistics Agency (3)	constituency
Used to generate new spatial data			
Imagery_2014	High resolution satellite imagery	2014-2016 Maxar (DigitalGlobe) Quickbird imagery, 30cm (4)	Coordinate (2016 household)
Imagery_2004	High resolution satellite imagery	2004-2016 Maxar (DigitialGlobe) SPOT imagery, 40cm (4)	Coordinate (2001, 2006, 2011 household)
census_ea	2011 Census EA boundaries	2011 Namibia Statistics Agency (5)	EA
Spatial covariates			
ccilc_dst011_2012	Dist to land-cover: Cultivated terrestrial lands	2008-2012 GlobCover, 300m (6)	100m
ccilc_dst040_2012	Dist to land-cover: Woody / Trees	2008-2012 GlobCover, 300m (6)	100m
ccilc_dst130_2012	Dist to land-cover: Shrubs	2008-2012 GlobCover, 300m (6)	100m
ccilc_dst140_2012	Dist to land-cover: Herbaceous	2008-2012 GlobCover, 300m (6)	100m
ccilc_dst150_2012	Dist to land-cover: Other vegetation	2008-2012 GlobCover, 300m (6)	100m
ccilc_dst190_2012	Dist to land-cover: Urban	2008-2012 GlobCover, 300m (6)	100m
ccilc_dst200_2012	Dist to land-cover: Bare	2008-2012 GlobCover, 300m (6)	100m
cciwat_dst	Dist to water bodies	2000 OSM (7)	100m
dmsp_2011	Night-time lights intensity	2012 Suomi VIIRS, 500m (8)	100m
gpw4coast_dst	Dist to coastline	GPWv4, 1km (9)	100m
osmint_dst	Dist to road intersections	2000 OSM (7)	100m
osmriv_dst	Dist to major water ways	2000 OSM (7)	100m
slope	Slope	2000 HydroSHEDS, 100m (10)	100m
topo	Elevation	2000 HydroSHEDS, 100m (10)	100m
tt50k_2000	Travel time to populated places	2000 JRC-EC (11)	100m
urbpx_prp_1_2012	Proportion of urban pixels with 1 cell radius	2009 Modis (12); Global Human Settlement City Model, 1km (13)	100m
hfacilities_dst	Dist to health centre or hospital	2001 UN-OCHA (14)	100m
schools_dst	Dist to primary/secondary school	2001 UN-OCHA (15)	100m
npp_2012	Annual net primary productivity	2010 MODIS, 1km (16)	100m

1. (ICF International, 2020)
2. (NSA, 2013)
3. (NSA, 2011b)
4. (Maxar, 2019)
5. (NSA, 2011a)
6. (European Space Agency, 2012)
7. (OpenStreetMap contributors, 2000)
8. (NOAA, 2012)
9. (CIESIN, 2018)
10. (Lehner, Verdin and Jarvis, 2006)
11. (Nelson, 2008)
12. (Schneider, Friedl and Potere, 2009, 2010)
13. (European Commission, 2017)
14. (UN-OCHA-ROSA, 2001b)
15. (UN-OCHA-ROSA, 2001a)
16. (Steven W. *et al.*, 2004)

3.4.1.4 Phase A: Methods and results

In phase A, I predicted the spatial distribution of various household types in Khomas using DHS data, spatial covariates, and visual inspection of satellite imagery. The output was a probability surface for each household type. Four steps were taken in phase A as follows, and are outlined in Figure 9.

(1) In step A1, I used all variables common between the 2013 DHS and 2011 census to define household types. These included urban/rural location, toilet type, water source, number of sleeping rooms, floor material, cooking fuel type, education level of each household member, and age and sex of each household member. Using the 2013 DHS data of 931 households in Khomas, I calculated eight demographic and social characteristics summarized to 53 PSU locations (i.e., clusters). Characteristics were coded at the household-level as binary, and at the PSU-level as percentages: urban, improved toilet, improved water source, sufficient sleeping space, durable structure, non-solid fuel for cooking, whether the head of household had any formal education, and whether there were any children under age five. “Improved” toilet and water source were defined according to the 2013 DHS which follows World Health Organization guidelines (WHO, 2012).

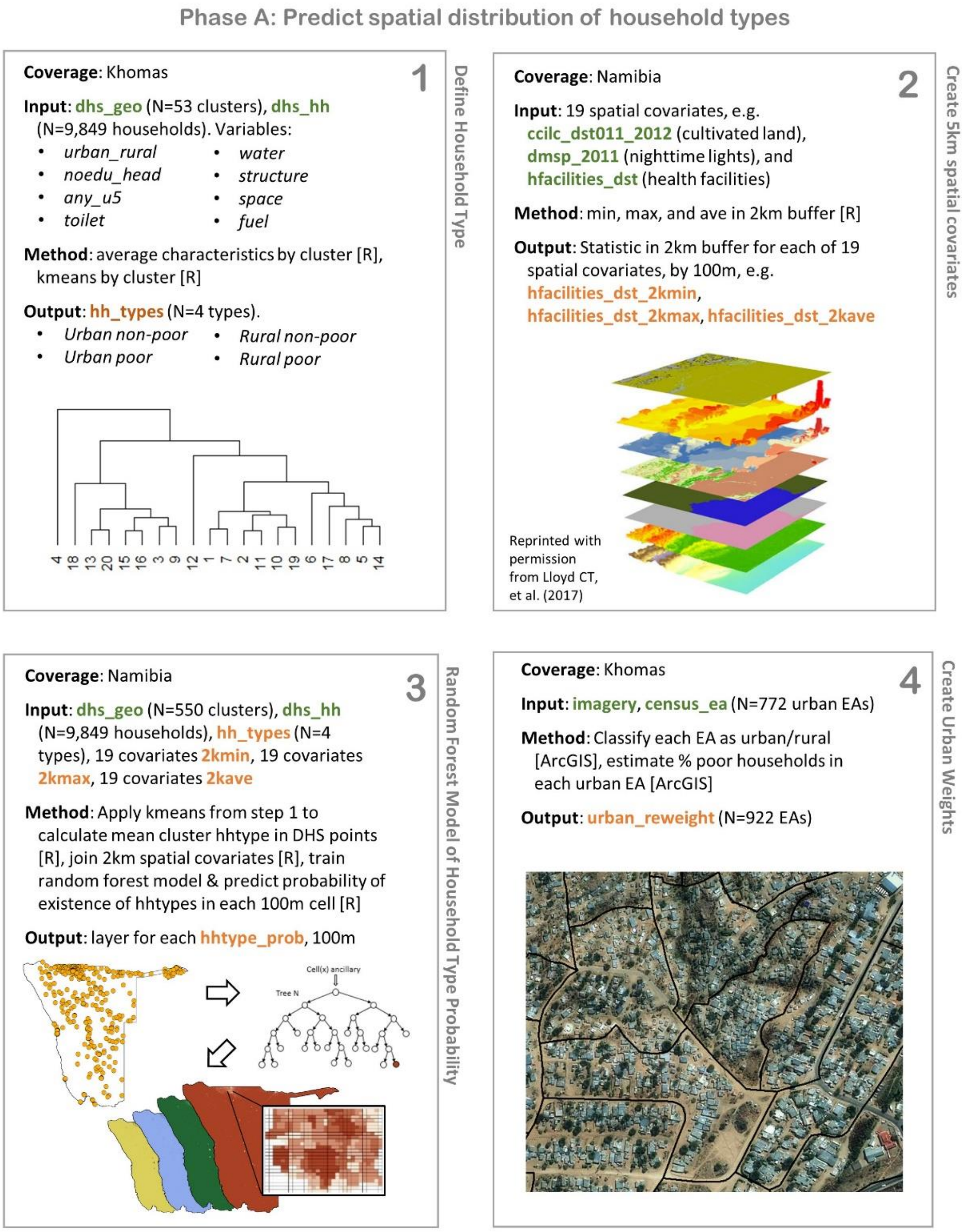


Figure 9. Phase A simulation steps to create household types and probability surfaces for each household type location in Khomas, Namibia

Adopted with permission from (Thomson, Kools and Jochem, 2018). Orange text is an original dataset, green text is a created dataset

I used k-means multivariate statistical methods (Groves *et al.*, 2009) to generate an arbitrary large number of clusters (20 potential household types) from the eight demographic and social variables common to the 2011 Census and 2013 DHS (Hartigan and Wong, 1979). Use of a dendrogram allowed me to display the Euclidean distance between each two child clusters and their combined parent cluster (Figure 10). I chose to use the first four clusters (cut off = 1.0) to define household types because the distinctions were easy to interpret as urban poor, urban non-poor, rural poor, and rural non-poor (Table 5). I considered use of the first six clusters (cut off = 0.75), but the interpretation of household type was unclear, and the statistical distinction between six and seven clusters (observed in the dendrogram) were minor (Appendix A).

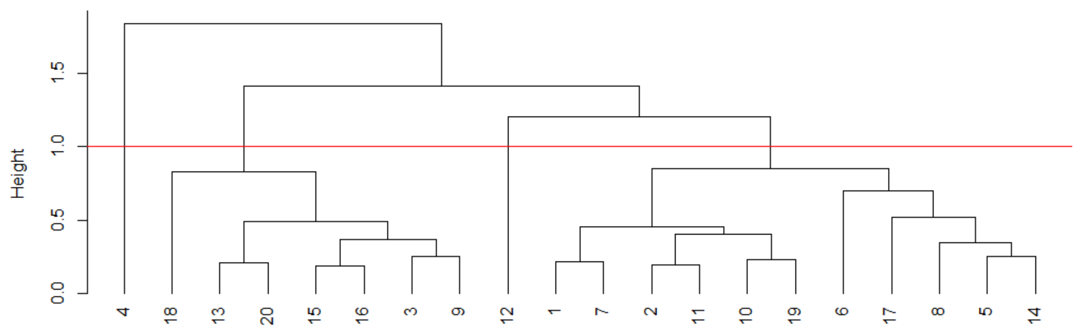


Figure 10. Dendrogram of 2013 DHS household types in Khomas, Namibia

Table 5. Average prevalence of variables for each k-means household type in Khomas, Namibia

Red indicates that value is 0.05+ above the Khomas average (less desirable), and green indicates that value is 0.05+ below the Khomas average (desirable)

Cluster	**rural**	**noedu**	**any_u5**	**unimp. toilet**	**unimp. water**	**inadeq. structure**	**limited space**	**solid fuel**	**Household type label**
Type 1	0.00	0.02	0.30	0.30	0.01	0.08	0.10	0.01	Urban non-poor
Type 2	0.00	0.10	0.35	0.96	0.01	0.54	0.09	0.24	Urban poor
Type 3	1.00	0.33	0.40	1.00	0.13	0.53	0.14	1.00	Rural poor
Type 4	1.00	0.04	0.21	0.43	0.00	0.08	0.00	0.00	Rural non-poor
Khomas	0.06	0.05	0.31	0.50	0.01	0.21	0.09	0.09	

(2) In step A2, I converted the 19 spatial covariates (from Table 4) to 100 meter by 100 meter grid cells, the scale at which I would later aggregate the simulated "real" population as a reference for gridded population data accuracy assessments. For each cell, I calculated two kilometre average, minimum, and maximum values for each covariate. I used a two kilometre moving window average because the DHS displaces urban PSUs up to this distance (Burgert, Zachary and Colston, 2013; Perez-Heydrich *et al.*, 2016), and the vast majority of households in Khomas are urban. These urban probability surfaces were further improved manually in step A4.

(3) In step A3, I calculated the most common household type (as defined in step A1) in each of the 550 PSUs in the Namibia 2013 DHS, and extracted the two kilometre average covariate values (from step A2) for each PSU. The distribution of average household type in DHS clusters was: 185 (34%) Type1 - urban non-poor, 82 (15%) Type 2 - urban poor, 249 (45%) Type 3 - rural poor, and 34 (6%) Type 4 - rural non-poor. I used the 550 cluster household types as training data, and the average covariate values across Namibia in a Random Forest machine classification model to predict a probability surface for each household type in each 100 meter by 100 meter cell in Namibia (Breiman, 2001).

The model performed well for Type 1 - urban non-poor and Type 3 - rural poor household types, but poorly for the other two household types as indicated by model classification error in

Table 6. Covariate performance in Random Forest models is measured with importance scores, summarising average variance explained over all regression trees. Table 7 reveals low importance scores across all covariates for Type 2 - urban poor and Type 4 - rural non-poor areas, but high importance scores for multiple covariates associated with Type 1 - urban non-poor areas (i.e. mean distance to intersections and schools, and mean and maximum proportion of urban pixels in a two kilometre buffer) and Type 3 - rural poor areas (i.e. mean distance to intersections, minimum distance to cultivated land, minimum elevation, and mean proportion of urban pixels in a two kilometre buffer). While it was not a surprise that covariate values generalized to two kilometre buffers did a poor job distinguishing household types within urban and rural areas, I was particularly concerned about the spatial allocation of poor and non-poor households within Windhoek. Households are highly segregated by socioeconomic class in cities around the world (UN-Habitat, 2016), and visual inspection of satellite imagery indicated that there was substantial socioeconomic segregation in Windhoek as well.

Table 6. Random Forest confusion matrix for average household type in 550 DHS clusters in the Khomas, Namibia simulation

	Type 1	Type 2	Type 3	Type 4	Classification Error
Type 1	158	23	3	1	0.14594595
Type 2	40	34	7	1	0.58536585
Type 3	8	3	230	8	0.07630522
Type 4	4	0	22	8	0.76470588

Table 7. Random Forest importance scores for average household type in 550 DHS clusters in the Khomas, Namibia simulation

Covariate	Type 1	Type 2	Type 3	Type 4
osmint_dst_mean	25.161662	8.89861202	20.091511	2.4142762
urbpx_prp_1_2012_mean	23.69525	15.415276	21.666851	2.6154466
schools_dst_mean	20.53483	3.86763197	12.046593	2.6770044
urbpx_prp_1_2012_max	14.714099	8.58676466	14.164805	1.5658259
NPP_2012_mean	13.973209	5.20754873	12.809764	1.9270368
ccilc_dst011_2012_min	12.940781	4.2685531	19.327524	3.9007588
NPP_2012_max	10.395984	5.25913327	16.285916	2.2096814
topo_min	9.72906	8.36047528	18.804455	0.9571754
hfacilities_dst_min	9.64691	15.37554307	12.211315	1.6536184
ccilc_dst190_2012_min	8.977913	11.49792198	8.052405	6.8093129
ccilc_dst200_2012_min	7.488454	7.1721323	8.971329	-2.7270707
ccilc_dst130_2012_max	6.986184	10.44480083	-1.504759	-0.8511626
osmriv_dst_mean	5.028809	9.27800633	8.447977	1.5117583
osmroa_dst_min	4.725465	-0.06461723	1.951167	1.7723995

(4) In step A4, I manually assigned urban poor and urban non-poor probability weights to census EAs to reduce the classification error of Type 2 (i.e., urban poor) households. Before beginning this process, I split several large census EAs located around the periphery of Windhoek to create new EAs for areas that had undergone urban expansion since the 2011 census boundaries were drawn (Figure 11). This increased the number of EAs in Khomas from 920 in the 2011 census to 922 for this 2016 simulation. I used the 922 EAs for this simulation, but in later analyses, I dissolved the boundaries of two small, narrow EAs in Windhoek city centre (Figure 11).

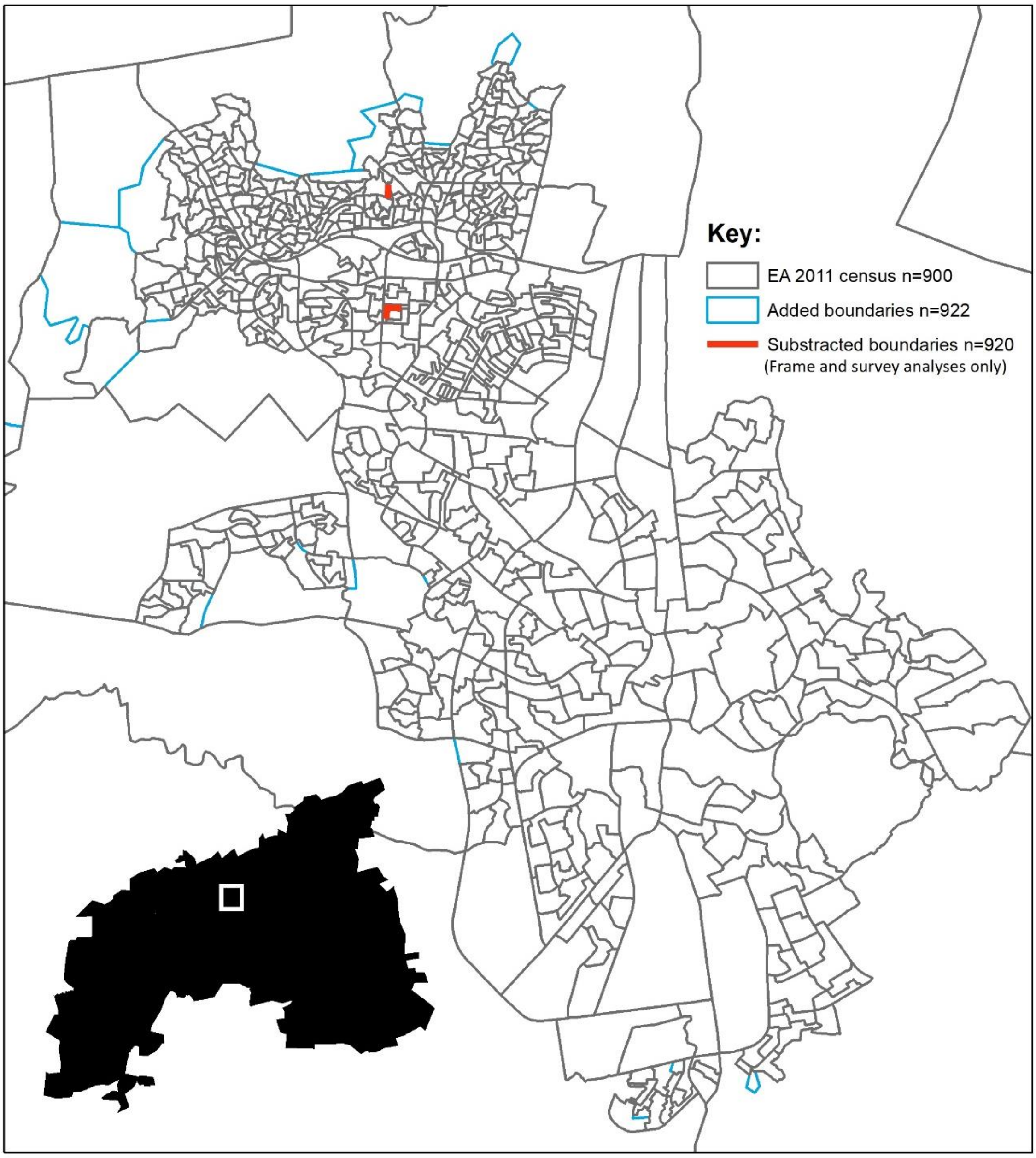


Figure 11. Modifications to 2011 census EA boundaries in and around Windhoek, Namibia to reflect 2016 urban expansion for population simulation

I manually created urban poor/non-poor probability weights for each of the 922 EAs based on visual inspection of satellite imagery, such that the probabilities summed to one within each EA. I considered areas comprised of organized larger buildings surrounded by green space to be non-poor, and areas characterised by disorganized smaller buildings located among earthen paths to be poor. Probability weights reflected my judgement of the proportion of buildings (and households) located in poor/non-poor areas within a given EA. Rural EAs had a null probability in this step. I multiplied the predicted household type probability surfaces created in step A3 by these manually created EA-level poor/non-poor household probability weights to create final 100 meter by 100 meter household type probability surfaces (Figure 12).

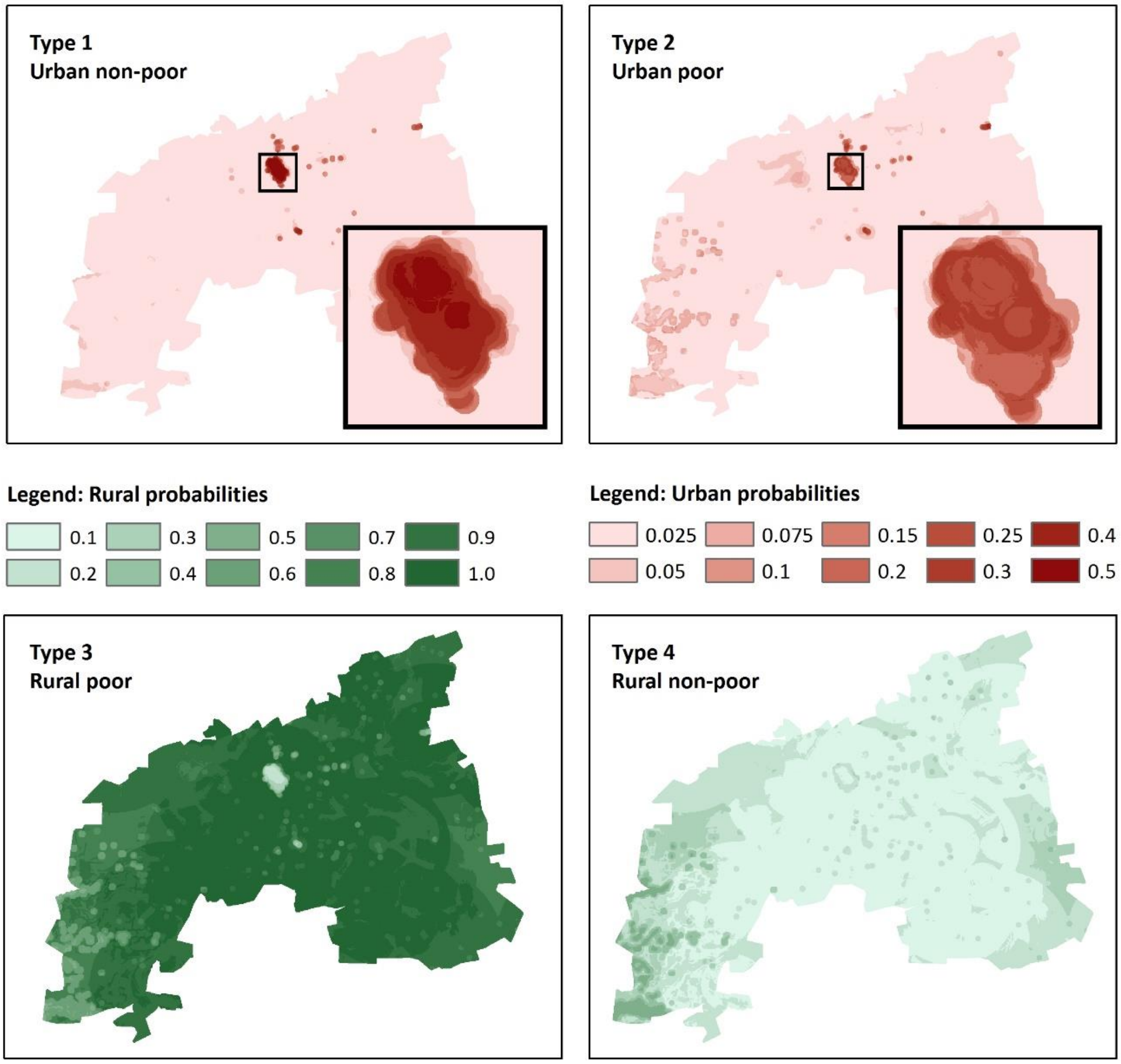


Figure 12. Household type probability surfaces from step A4 in Khomas, Namibia population simulation

3.4.1.5 Phase B: Methods and results

Phase B was comprised of steps 5 through 7 to geo-locate the synthetic population to realistic building locations (Figure 13).

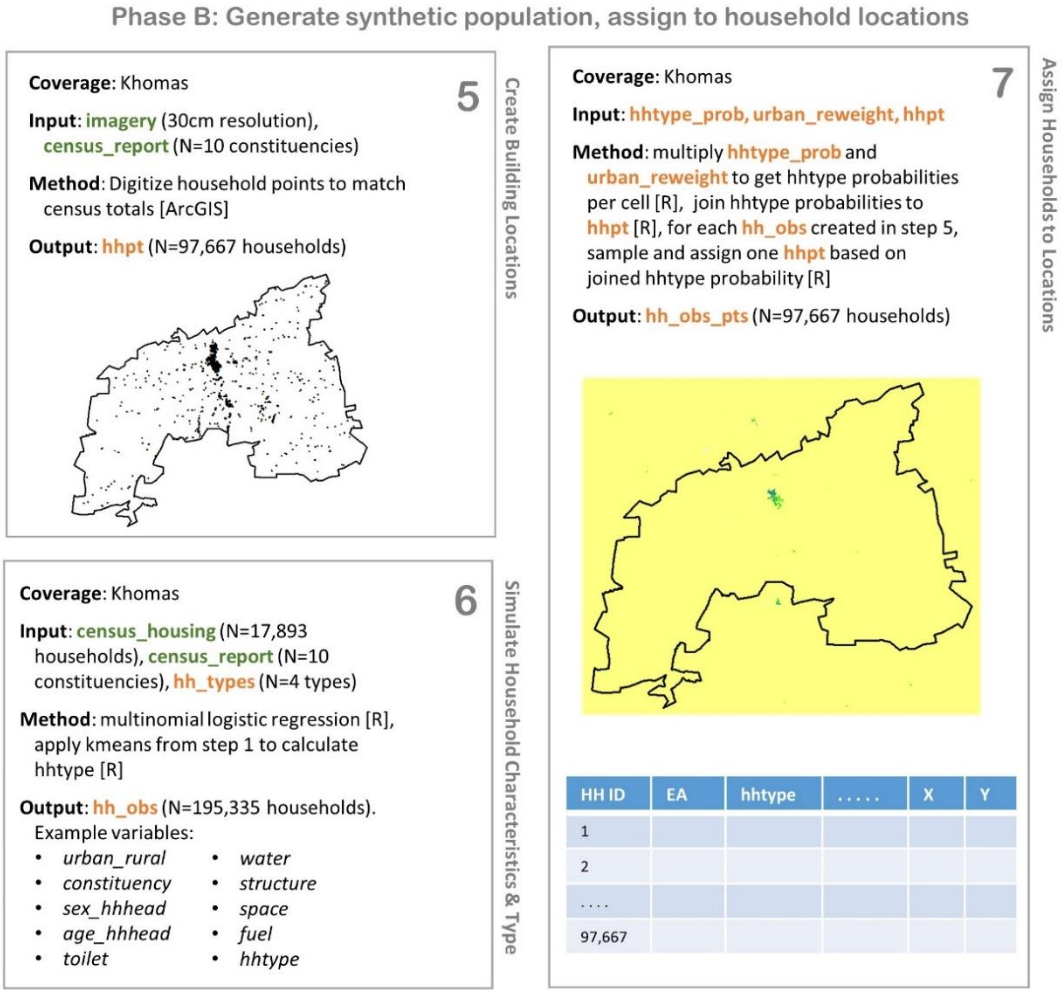


Figure 13. Phase B simulation steps to generate a synthetic population and assign simulated households to realistic geo-locations locations in Khomas, Namibia

Adopted with permission from (Thomson, Kools and Jochem, 2018). Orange text is an original dataset, green text is a created dataset.

(5) In step B5, I manually digitized realistic household locations across Khomas using 2014-2016 high-resolution (30cm) Quickbird imagery in ArcGIS 10.5. Subjective judgement was required for this process, for example, deciding not to digitize buildings on main streets in densely populated areas where non-residential shops and offices seemed likely. I also chose not to digitize main buildings in industrial compounds because I assumed they were unoccupied factories or warehouses; however, in larger compounds I digitized buildings that appeared to be residential barracks. In areas of dense settlement, I duplicated some points to represent more than one household in the same building. A total of 97,667 points (i.e., households) were digitized in 2016 in Khomas. As a benchmark, I exported points to Google Earth and used 2011 Maxar and SPOT (40cm) imagery to identify buildings that were missing in 2011, and ensured that the reduced number of points matched constituency household counts in the 2011 census (NSA, 2011b).

(6) In step B6, I simulated a population of realistic households in Khomas using iterative proportional fitting (IPF) with combinatorial optimisation in the R *simPop* package (Alfons *et al.*, 2011; Templ *et al.*, 2017) (Table 8). IPF starts by defining a basic household structure to ensure the synthetic population is realistic. I defined household structure with household size, urban/rural residence, age of household head, and sex of household head at the household-level, and age, sex, and relationship (to household head) at the individual-level. Inputs to the model were the 2011 Population and Housing Census 20% microdata sample, as well as urban and rural household sizes, and constituency population by age, sex, and relationship based on the 2011 census report (NSA, 2011b). The IPF model selects random samples of records from the microdata with replacement until each of the household structure targets per constituency are met.

Table 8. Iterative proportional fitting of household structure in Khomas, Namibia simulation by constituency

	Tobias Hainyeko	Katutura Central	Katutura East	Khomasdal North	Soweto	Samora Machel	Windhoek East	Windhoek Rural	Windhoek West	Moses Garoëb
N	60553	30868	24078	60465	19570	80036	27309	30028	62588	62807
HH Size										
Average	5.49	5.49	5.49	5.49	5.49	5.49	5.49	5.49	5.49	5.49
Residence										
Urban	100%	100%	100%	100%	100%	100%	100%	26%	100%	100%
Rural	0%	0%	0%	0%	0%	0%	0%	74%	0%	0%
Relationship										
Head	27%	21%	20%	24%	22%	26%	34%	30%	28%	30%
Spouse	10%	6%	5%	9%	6%	8%	18%	13%	13%	9%
Child	26%	27%	27%	31%	25%	27%	28%	28%	29%	23%
Grandchild	4%	8%	12%	4%	10%	6%	1%	7%	2%	5%
Extended	29%	31%	29%	26%	31%	28%	12%	14%	20%	29%
Other	5%	8%	7%	6%	5%	5%	8%	7%	8%	5%
Sex										
Female	45%	55%	56%	53%	53%	52%	51%	46%	53%	47%
Male	55%	45%	44%	47%	47%	48%	49%	54%	47%	53%
Age										
<1	4%	2%	3%	3%	2%	3%	2%	3%	2%	4%
1 - 4	9%	8%	9%	8%	7%	9%	7%	9%	7%	9%
5 - 9	9%	10%	10%	9%	9%	8%	6%	10%	7%	8%
10 - 14	8%	10%	10%	10%	9%	9%	6%	10%	8%	6%
15 - 19	8%	11%	11%	11%	11%	10%	8%	9%	11%	7%
20 - 24	15%	12%	13%	14%	17%	15%	8%	9%	15%	14%
25 - 29	14%	12%	10%	10%	12%	14%	9%	8%	10%	15%
30 - 34	11%	10%	8%	9%	9%	11%	9%	7%	9%	13%
35 - 39	9%	7%	7%	8%	6%	7%	9%	7%	7%	11%
40 - 44	6%	5%	5%	6%	4%	5%	9%	7%	6%	6%
45 - 49	4%	4%	4%	5%	3%	4%	6%	5%	5%	4%
50 - 54	2%	3%	3%	3%	4%	2%	6%	5%	4%	2%
55 - 59	1%	2%	2%	2%	3%	2%	5%	3%	3%	1%
60 - 64	1%	1%	2%	1%	1%	1%	3%	3%	2%	1%
65 - 74	0%	1%	2%	1%	1%	1%	5%	4%	2%	0%
75+	0%	1%	1%	1%	0%	0%	2%	2%	1%	0%

Next, in the R *simPop* package, I added household and individual characteristics present in the 20% microdata census dataset (toilet, water, structure, space, fuel, education) to the simulated dataset using a multinomial logistic regression technique and conditional annealing (Table 9) (Alfons *et al.*, 2011; Templ *et al.*, 2017). This treated age, sex, relationship, household size, and urban/rural residence as predictors, and each of the household characteristic as a conditional outcome.

Table 9. Multinomial logistic regression output of household characteristics in Khomas, Namibia simulation by constituency

	Tobias Hainyeko	Katutura Central	Katutura East	Khomasdal North	Soweto	Samora Machel	Windhoek East	Windhoek Rural	Windhoek West	Moses Garoëb
N (individuals)	60553	30868	24078	60465	19570	80036	27309	30028	62588	62807
Water										
Improved	100%	100%	100%	100%	100%	100%	100%	96%	100%	100%
Unimproved	0%	0%	0%	0%	0%	0%	0%	4%	0%	0%
Toilet										
Improved	25%	58%	67%	76%	69%	44%	97%	52%	94%	24%
Unimproved	75%	42%	33%	24%	31%	56%	3%	48%	6%	76%
Floor										
Durable	44%	97%	99%	88%	96%	72%	96%	80%	98%	44%
Non-durable	56%	3%	1%	12%	4%	28%	4%	20%	2%	56%
Space										
Adequate	81%	64%	64%	78%	74%	74%	96%	75%	93%	81%
Inadequate	19%	36%	36%	22%	26%	26%	4%	25%	7%	19%
Fuel										
Non-solid	87%	99%	97%	93%	99%	94%	100%	50%	100%	92%
Solid	13%	1%	3%	7%	1%	6%	0%	50%	0%	8%
HH Head Education										
No formal	24%	20%	21%	18%	16%	21%	14%	30%	14%	24%
Some primary	22%	20%	19%	19%	17%	18%	10%	24%	12%	20%
Primary	37%	38%	35%	32%	32%	36%	14%	28%	18%	38%
Secondary	15%	19%	20%	22%	26%	21%	33%	12%	32%	18%
Tertiary	2%	3%	5%	9%	8%	4%	29%	6%	24%	1%

I confirmed that there were not major differences between the distributions of characteristics in the 20% microdata and simulated dataset (Table 10). Confident that the simulated household and individual characteristics were realistic, I calculated the most likely household type for each household based on variable factor weights created in the k-means analysis in step A1.

Table 10. Differences in percent distribution between (weighted) Khomas census 20% microdata sample and simulated dataset in Khomas, Namibia

Constituency	**Diff**
Tobias Hainyeko	-0.002
Katutura Central	0.001
Katutura East	-0.001
Khomasdal North	0.000
Soweto	-0.001
Samora Machel	-0.002
Windhoek East	0.000
Windhoek Rural	-0.001
Windhoek West	0.002
Moses Garoëb	0.001

Residence	**Diff**
Urban	0.000
Rural	0.000

Water	**Diff**
Improved	-0.002
Unimproved	0.002

Toilet	**Diff**
Improved	0.000
Unimproved	0.000

Floor	**Diff**
Durable	-0.001
Non-durable	0.001

Space	**Diff**
Adequate	0.002
Inadequate	-0.002

Fuel	**Diff**
Non-solid	-0.001
Solid	0.001

Household size	**Diff**
1	0.000
2	0.000
3	0.000
4	0.000
5	0.000
6	0.000
7	0.000
8	0.000
9	0.000
10	0.000
11	0.000
12	0.000
13	0.000
14	0.000
15	0.000
16	0.000
17	0.000
18	0.000
19	0.000
20	0.000
21	0.000
24	0.000
29	0.000
30	0.000
31	0.000
32	0.000
33	0.000

Sex	**Diff**
Female	0.002
Male	-0.002

Relationship	**Diff**
Head	0.000
Spouse	0.000
Child	-0.001
Grandchild	0.000
Extended	0.001
other	0.000

Age	**Diff**
<1	0.000
1 - 4	0.000
5 - 9	-0.001
10 - 14	0.000
15 - 19	0.000
20 - 24	0.000
25 - 29	0.000
30 - 34	0.000
35 - 39	0.000
40 - 44	0.000
45 - 49	0.000
50 - 54	0.000
55 - 59	0.000
60 - 64	0.000
65 - 74	0.000
75+	0.000

Education	**Diff**
No formal	0.000
Some primary	-0.001
Primary	-0.001
Secondary	0.000
Tertiary	0.002

Finally, I scaled up the number of simulated observations to ensure there were enough simulated households to assign to the digitized 2016 coordinates created in step B5. The 2011 microdata sample was provided with a weight of approximately five for each observation to scale the 20% microdata sample to the total population in 2011. I calibrated the simulation to create an extra 20% of households to ensure there were enough simulated households to assign to 2016 point locations; left over simulated households were discarded in step B7. This resulted in 122,079 simulated households in Khomas before assignment to point locations.

(7) In step B7, I joined the re-weighted household type probabilities created in step A4 to the household latitude-longitude coordinates created in step B5. For each latitude-longitude coordinate created for 2016 household point locations, I randomly sampled a simulated household created in step B6 from the corresponding constituency and urban/rural strata based on the probabilities of household types at each coordinate. I repeated assignment of simulated households to coordinate point locations until all coordinates were assigned a simulated household, and then discarded the extra unassigned simulated households for a total of 97,667 simulated households located at realistic coordinate locations in Khomas for 2016.

3.4.1.6 Phase C: Methods and results

Phase C consisted of just one step, to predict population outcomes in the synthetic population based on information in the 2013 DHS (Figure 14).

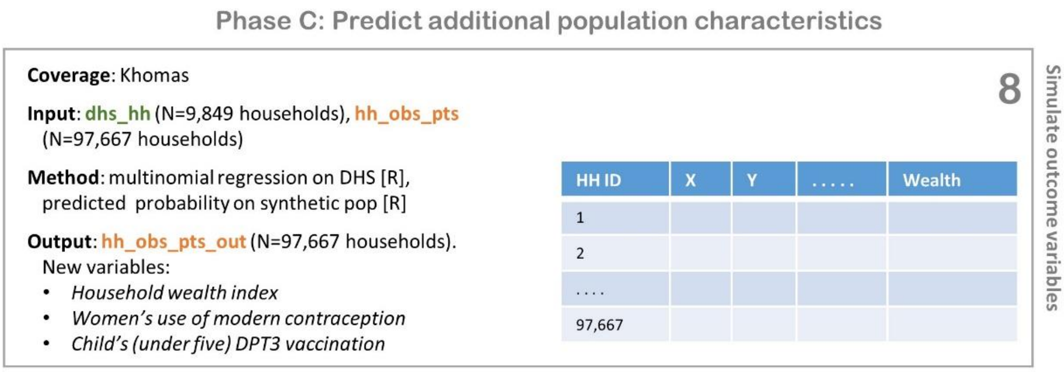


Figure 14. Phase C simulation steps to generate outcome characteristics in the synthetic population in Khomas, Namibia

Adopted with permission from (Thomson, Kools and Jochem, 2018). Orange text is an original dataset, green text is a created dataset.

(8) In step C8, I used the 2013 DHS records in Khomas (n=931 households) to develop multinomial models in R to simulate individual and household outcomes that are typically collected in a household survey. Based on the recommendations of survey experts Dr. Felicity Cutts and Mr. Dale Rhoda, three outcomes were chosen to represent different prevalence levels and patterns of dispersion in the population, as well as clustering of children, in women, in households: household wealth quintile (five ordinal categories), woman's use of modern contraception (binary in women age 15 to 49), and child's receipt of 3rd DPT vaccination (binary in children under five). Dr. Kools leveraged these outcomes in the Oshikoto simulation to compare

survey sample designs (Kools, 2018). I simulated the same outcomes in step C8 in Khomas to test sampling accuracy later in the thesis. As in the Oshikoto simulation (Thomson, Kools and Jochem, 2018), I used a multinomial model to calculate associations between each outcome and household-level covariates in the 2013 DHS dataset (Table 11). I stored the coefficients from the 2013 DHS models and applied them to the simulated dataset to predict wealth quintile, modern contraceptive use, and receipt of 3rd DTP vaccine for each household, woman 15 to 49, and child under five, respectively.

Table 11. Multinomial model coefficients and fit statistics for three outcomes in the 2013 DHS for Khomas, Namibia

Predictor	Household wealth quintile (ref=poorest)				Women 15-49 use of modern contraception	Child <5 DPT3 vaccination coverage
	poorer	middle	richer	richest		
Rural	0.479	0.773*	2.299***	2.061***	-0.227**	2.334***
HH Head						
15-29	(ref.)	(ref.)	(ref.)	(ref.)		
30-49	-11.595***	-11.222***	-11.581***	-10.890***		
50+	-9.957***	-9.171***	-8.901***	-7.715***		
HH Head Female	1.003***	0.778**	0.929**	0.333		
Age						
15 – 19					-1.290***	
20 – 24					-0.111**	
25 - 29					0.208***	
30 – 34					(ref.)	
35 – 39					0.030	
40 - 44					0.123**	
45 - 49					-0.023	
Child age 1 – 4						0.795***
Female						-0.188***
HH Head						
No education	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Some primary	0.133	-0.133	0.121	0.166	0.562***	0.680***
Primary	1.459***	2.243***	2.401***	3.216***	-0.038	0.447***
Secondary	0.466	1.651***	2.675***	4.092***	0.023	0.258
Tertiary	4.844***	6.455***	7.491***	9.515***	-0.259***	0.667***
Water Unimproved	-1.262*	0.429	-106.655	-0.169	-0.023	11.129
Toilet Unimproved	-23.935***	-26.157***	-28.908***	-30.603***	-0.018	0.021
Space Inadequate	-0.771**	-1.652***	-0.292	-1.216***	0.028	0.293***
Floor Non-durable	-21.756***	-22.962***	-24.338***	-26.003***	0.297***	0.748***
Fuel Solid	-19.316***	-20.937***	-23.301***	-105.303***	-0.197**	-0.621***
Constant	77.205***	80.003***	82.729***	82.498***	0.446***	-0.250
AIC	30,400				27,470	6,344

Note: *p<0.1; **p<0.05; ***p<0.01

To evaluate the realism of the simulated outcomes, I aggregated DHS records to PSU, and aggregated simulated records to EA, then compared the inter-cluster correlation (ICC) and distributions of household and individual characteristics and outcomes. The household characteristics and outcomes – average household size, average wealth quintile, percent with improved toilet, percent with improved water source, percent with sufficient sleeping space, and percent with durable floors – appeared to be consistent between the DHS and the simulated population in terms of both distribution and ICC (Figure 15). Individual characteristics and outcomes – average age, percent female, prevalence of modern contraceptive use among women, and DPT3 vaccine coverage among children under five - were less consistent, and more heaped around the mean in the simulated dataset, likely due to more observations in the simulated dataset (892 EAs) compared to the 2013 DHS dataset (53 clusters) (Figure 16). For this reason, I only analyse household-level covariates in later analyses.

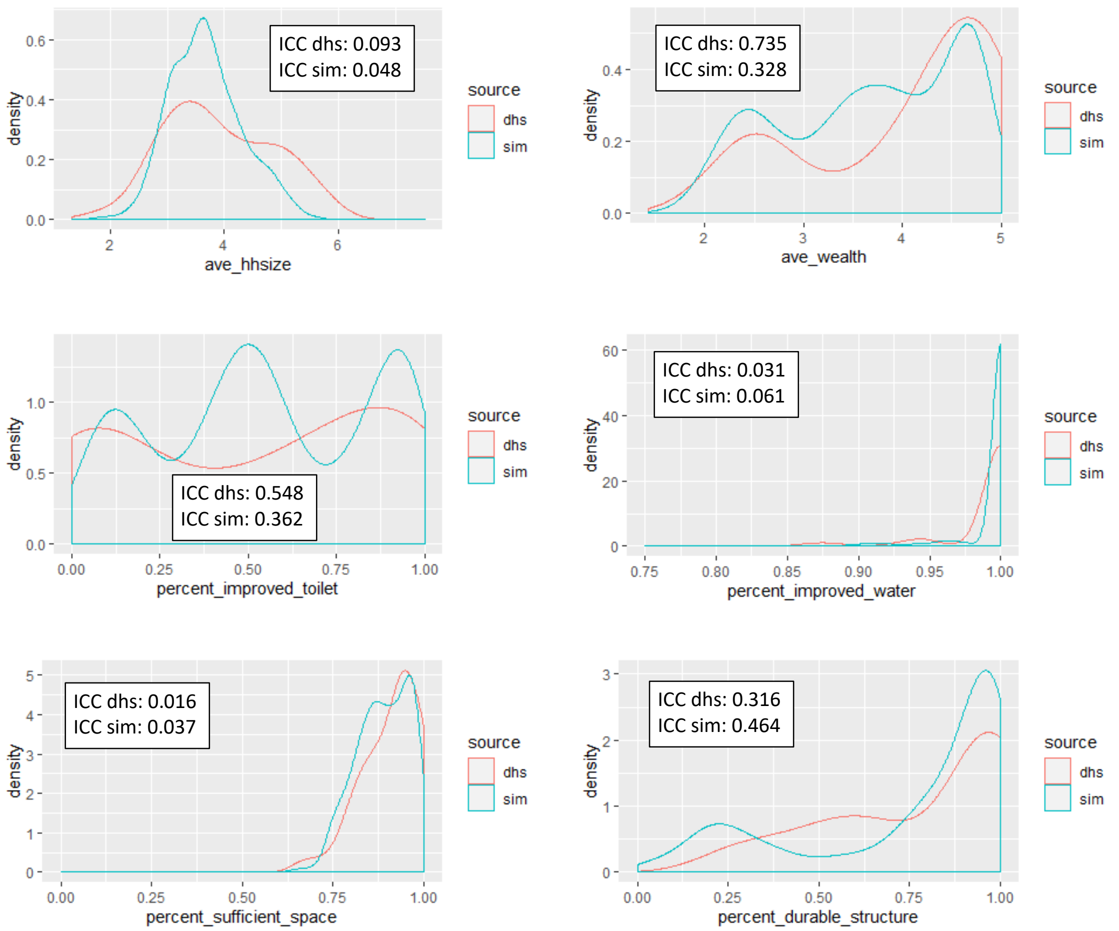


Figure 15. Comparison of household outcomes by 2013 Namibia DHS cluster (Khomas region only) and simulated population EA in Khomas, Namibia

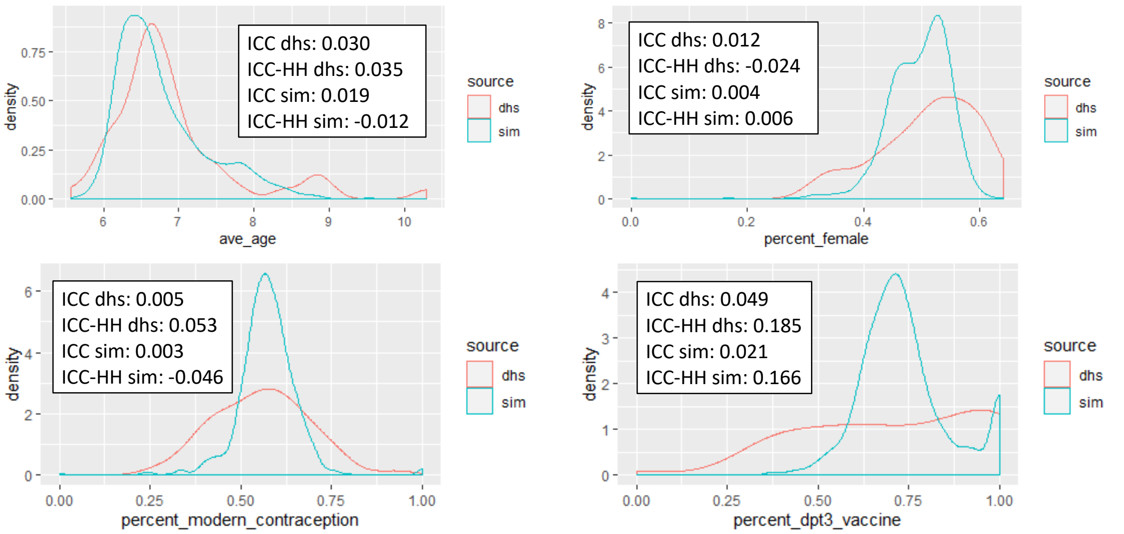


Figure 16. Comparison of individual outcomes by 2013 Namibia DHS cluster (Khomas region only) and simulated population EA in Khomas, Namibia

3.5 Simulating 16 censuses with realistic outdatedness and inaccuracies

Satisfied that I had simulated a realistic “true” population in Khomas in 2016, I set out to simulate several realistic census datasets with varying degrees of outdatedness and inaccuracy. The simulation of realistically imperfect census datasets from the “true” population are important to generate realistic gridded population datasets for later analyses in chapters 4 and 5. This is particularly relevant because survey practitioners often turn to gridded population sampling when they feel that the standard census frame is too outdated or inaccurate to sample directly, with an implied assumption that samples selected from gridded population datasets derived from outdated or inaccurate surveys will be more accurate.

One in four countries located in Africa, Asia, Oceania and Latin America has not had a census in 15 or more years (UNSD, 2019) or is a recipient of long-term humanitarian aid due to protracted or recurring crises and population displacement (Development Initiatives, 2018) (Figure 17). Given increasing levels of mobility and urbanization described in Chapter 2, censuses become outdated faster now than in the past, especially in urban populations (UN-Habitat, 2016). Even in LMICs with well-conducted, routine national censuses, population distributions change substantially over a decade. Population redistribution becomes a greater problem as censuses age, and is especially likely in countries where conflict, natural disasters, or economic distress have placed strong pressures on the population to relocate. Often, major political crisis or weak government are reasons for both census delays and large-scale population movements, resulting in the least amount of data in countries with the greatest data need. These countries include Afghanistan (last census 1979), DR Congo (last census 1984), Yemen (last census 1986), Somalia (last census 1987), and Iraq (last census 1997) (UNSD, 2019).

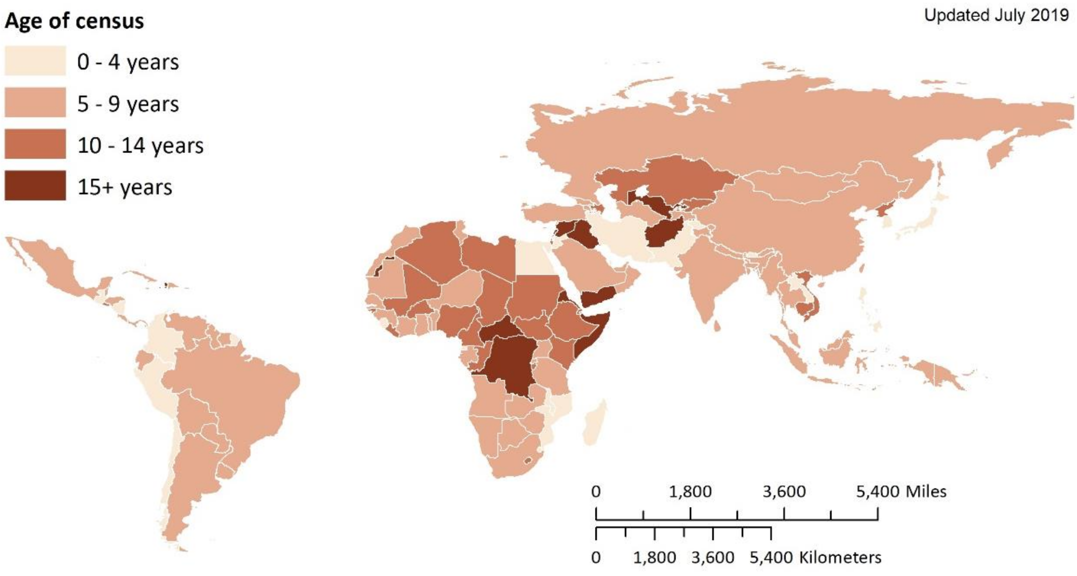


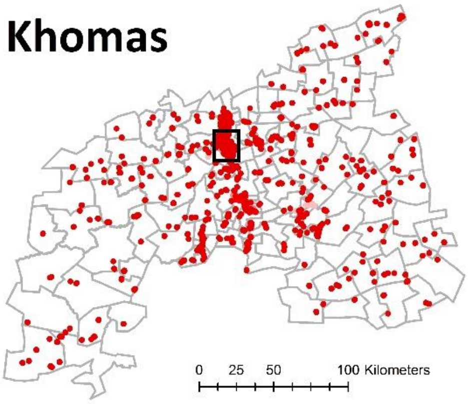
Figure 17. Map of last census year in Africa, Asia, Latin America, and Oceania

Data from 2020 World Population and Housing Census Programme (UNSD, 2019)

As described above (section 3.3), very high rates of population redistribution have occurred in Namibia in the three decades since independence, with most of the population shift occurring from rural to urban areas. As Khomas is dominated by the Windhoek metropolitan area, the bulk of the population redistribution is characterized by in-migration from other regions in Namibia or nearby countries. Based on this, I decided to simulate 5-, 10-, and 15-year-old census datasets from the "real" 2016 simulated population. For simplicity, I assumed that households present in 2001 still existed in 2016 and maintained the same household and individual characteristics. This means, that households were only added to Khomas between 2001 and 2016, and I did not apply any method to change household composition or outcomes over time.

3.5.1 Simulating census outdatedness in Khomas

To simulate population outdatedness in Khomas, I imported the 2016 simulated population household point locations into Google Earth, and used the software's historical Maxar and SPOT imagery (40cm) to identify buildings that were not present in 2011, 2006, and 2001 (Figure 18). The oldest imagery available at 40cm resolution in Google Earth for Khomas was from 2004, so I used some judgement to flag buildings that looked recently built in 2004 (e.g., bare fresh soil or rooftops that look incomplete without nearby foliage) and assumed they were not present in 2001. During this exercise, I ensured that the number of household coordinates in each constituency matched the number of households reported in the 2001 and 2011 Population and Housing Census final reports to ensure that both patterns and degree of outdatedness were realistic (NSA, 2011b).



Constituency	hh_2016	hh_2011	hh_2006	hh_2001
Tobias Hainyeko	12756	12428	10486	8872
Katutura Central	5182	5096	4948	4072
Katutura East	3824	3756	3659	3165
Khomasdal	11684	10471	7302	5770
Soweto	3470	3377	3167	2553
Samora Machel	16718	13250	7573	6598
Windhoek East	7532	7089	6451	5620
Windhoek Rural	7256	6330	5415	4961
Windhoek West	13947	13837	12590	9991
Moses //Garoëb	15298	13804	10315	6978
Khomas	*97667*	*89438*	*71906*	*58580*

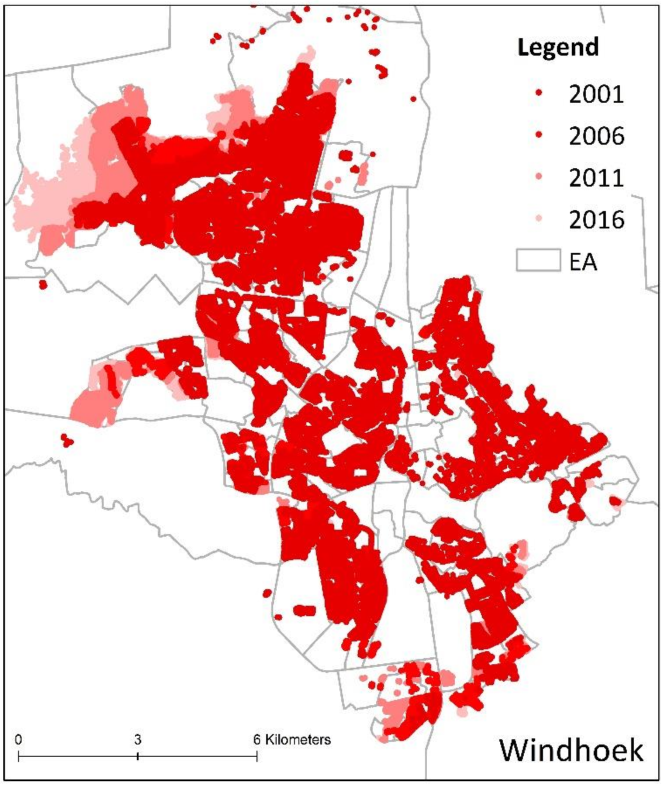


Figure 18. Household point locations in Khomas, Namibia classified by presence in years 2016, 2011, 2006, and 2001

3.5.2 Quantifying census inaccuracies in LMICs

Without a post-enumeration survey (PES) for the 2011 Namibia census, I was not able to derive realistic census inaccuracies from my data directly. Furthermore, I expected the Namibia 2011 census to be of high quality with few under-counts, which is why I chose to use Namibia's 2011 census as the foundation of this simulation. In order to quantify realistic ranges of census inaccuracies in LMICs, I turned to the scientific literature and PESs in other LMICs to (a) develop a household typology of sub-populations that are often under-counted in a LMIC census, and (b) perform a systematic literature review of LMIC census inaccuracy due to sub-population under-counts. Later in this chapter, I simulate several versions of an outdated-inaccurate census in Khomas with realistic low, medium, and high rates of inaccuracy.

3.5.2.1 Household typology

To identify meaningful sub-populations for the literature review of LMIC census under-counts, I started with UN census guidelines. The updated guidelines for the 2020 census round provide a typology and definitions of the following dwelling/household types: conventional dwellings, other housing units, guesthouses, institutions, and camps (UNSD, 2017). Outside of normal census activities, the guidelines recommend special enumerations of nomads, persons living in remote locations, and civilians/military personnel located outside of the country or traveling during the census (UNSD, 2017). For these special enumerations, the guidelines recommend that official registration data be used to count expatriates/travellers, and that one of the following methods be used to enumerate nomads and remote populations: meeting with tribal leaders, walking systematically across whole areas, or posting enumerators at known water points or camps (UNSD, 2017).

The guidelines, however, are unclear about the definitions of “roofless” and “secondary” homeless populations living in “other housing units”, and how these populations should be counted (UNSD, 2017). One might infer that “other housing units” include non-durable structures and mobile structures such as tents, and are likely located in slum-areas (Ezeh *et al.*, 2017; Lilford *et al.*, 2017; UN-Habitat, 2017). Furthermore, the guidelines do not mention methods to perform special enumerations of roofless homeless (street-sleepers) such as capture-recapture (Wright and Devine, 1992; Gurgel *et al.*, 2004; Stark *et al.*, 2017) or point-in-time counts (Peressini, Mcdonald and David, 2010; Troisi *et al.*, 2015) which have been used successfully in LMICs. For a clearer typology of the households and dwellings that comprise the urban poorest, I turned to Tipple and Speak (2005) who provide criteria for understanding and measuring homelessness in LMICs (Tipple and Speak, 2005). Combining the Tipple and Speak (2005) and UN 2020 census guidelines, I used the following categories to quantify LMIC census under-counts in the literature review:

Urban slum households

- Households in permanent/semi-permanent dwellings (e.g., shack or tent)
- Homeless, non-permanent households (e.g., couch surfer)
- Homeless, roofless households (e.g., street sleeper)
- Homeless, guest house dwellers (e.g., staying in temporary, paid accommodation)

Urban non-slum households

- Households in permanent dwellings

Rural households

- Households in permanent/semi-permanent dwellings (e.g., brick walls, traditional hut)
- Remote households (e.g., located in an area far from road access)
- Nomad households (e.g., move seasonally, non-permanent or semi-permanent dwelling)

3.5.2.2 Methods: Census inaccuracies in LMICs (Systematic Review)

To determine prevalence of census under-counts in LMICs, I reviewed census PESs and performed a systematic literature search of studies in LMICs. Only references reporting under-counts in sub-population in censuses conducted since 1990 were used.

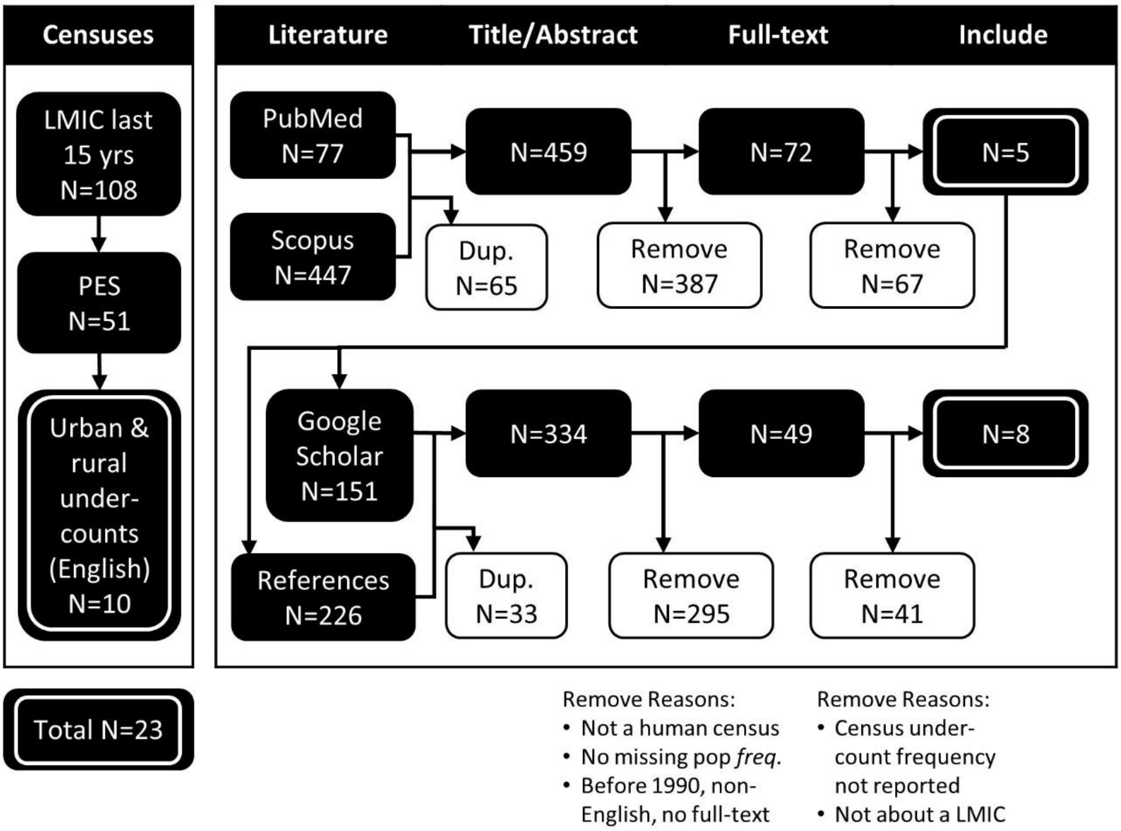
I performed the review of PESs in Google Chrome in English using the search phrase "census post enumeration survey [country name]", which means that PESs published in other languages might have been omitted. This search was conducted for the 108 LMICs listed on the UN Statistical Division Census Programme website and where a census had been conducted in the last 15 years (UNSD, 2019). Ten PESs from LMICs reported separate under-counts in urban and rural conventional households (Figure 19).

I then conducted a literature review in two waves. First, I searched PubMed and Scopus for peer-reviewed public health, economic, and related articles published from January 1, 1990 through February 28, 2017. The search was restricted to English-language articles about human population censuses that have taken place since 1990. The articles met all of the following criteria:

- The word "census" in the title or abstract
- One of the following terms in the title or abstract describing a problem of census coverage: listing, enumerat*, count, coverage, miss*
- One of the following terms in the title or abstract describing a census sub-population: nomad*, pastoral*, refugee, displaced, migrant, slum, poorest, unregistered, homeless, [street] sleeper, pavement [dweller], floating

The first wave of the literature search resulted in 459 unique articles, of which I screened all titles and abstracts. Of 72 potentially eligible articles from LMICs, I reviewed the full-text, and kept five which reported a census under-count (Figure 19).

In a second wave, I used Google Scholar to identify the top 20 "cited by" and top 20 suggested "related" articles for each of the five articles identified in the first wave of the review. I also searched all references of these articles, and then repeated the screening and full-text review. The second wave resulted in 334 unique articles, of which 49 had potentially relevant titles or abstracts. After a full-text review of these articles, I found eight reported census under-counts (Figure 19). Together, census under-counts in LMICs were estimated from 10 PESs, and 13 articles.



PES: Post-enumeration survey

Pubmed: Since 1990. (census*[Title/Abstract]) AND (listing*[Title/Abstract] OR enumerat*[Title/Abstract] OR count[Title/Abstract] OR coverage[Title/Abstract] OR miss*[Title/Abstract]) AND (nomad*[Title/Abstract] OR pastoral*[Title/Abstract] OR refugee[Title/Abstract] OR displaced[Title/Abstract] OR migrant[Title/Abstract] OR slum[Title/Abstract] OR poorest[Title/Abstract] OR unregistered[Title/Abstract] OR homeless[Title/Abstract] OR sleeper[Title/Abstract] OR pavement[Title/Abstract] OR floating[Title/Abstract])

Scopus: TITLE-ABS-KEY ((census OR censuses) AND (listing OR enumeration OR enumerate OR count OR coverage OR missing OR missed) AND (nomad OR nomadic OR pastoral OR pastoralist OR refugee OR displaced OR migrant OR slum OR poorest OR unregistered OR homeless OR sleeper OR pavement OR floating)) AND PUBYEAR > 1989

Google Scholar: Of the 5 articles kept and summarized in the first search ,performed an additional search in Google Scholar using the first 20 “cited by” and first 20 “related” articles.

References: Of the 5 articles, kept and summarized in the fist search, reviewed all references.

Figure 19. Search terms and process used in the census under-count literature review

3.5.2.3 Results: Census inaccuracies in LMICs (Systematic Review)

Results of the review are summarized in Table 12, and detailed in Appendix B. Census omissions are presented separately for rural, urban slum, and urban non-slum households, following the typology developed in Section 3.5.2.1. The literature search, however, resulted in several articles separately reporting census under-counts in settled versus mobile populations, or institutional populations specifically, thus it became important to track these population types across household types (rural, urban slum, urban non-slum).

Mobile populations in LMICs represent migrants, IDPs, refugees, homeless, and nomads or migratory pastoralists who may live in urban slum or rural areas. Institutional populations in LMICs refer to a wide range of populations with different living circumstances including refugee and IDP camps, workers' camps, homeless shelters, prisons, military barracks, religious orders, care homes, hospitals, schools, or university dormitories. While institutional populations can, technically, be located in any area type, I categorise them by the type of area in which they are most likely to be located. For example, urban IDP and refugee camps are most likely located in slum areas, and hospitals and care homes in cities are most likely to be located in non-slum areas. Some household-population type groupings do not exist (in any meaningful numbers), for example, mobile urban non-slum households living in permeant structures, and thus no estimates of under-counts are reported. Using this approach, I created 14 unique population-household groups and assigned each a range of census under-counts in LMICs (Table 12).

The greatest percentage of under-counts were among urban slum households; urban slum residents were roughly four times more likely to be excluded from a census than urban non-slum residents (Table 12). In urban slum areas, between 33% and 61% of the mobile population (not living in camps) were under-counted in recent LMIC censuses, and 5% to 59% of the non-mobile population living in permanent or semi-permanent dwellings were under-counted, compared to just 2% to 15% in non-slum urban areas (Table 12). The range of under-counts in urban non-slum areas was probably less than reported in Table 12, as these figures are from PES surveys which combined urban slum and non-slum populations.

According to several PESs, census under-counts in rural areas for residents of permanent and semi-permanent dwellings were as low as 2% and high as 13% among non-mobile populations (Table 12). Evidence suggests that mobile populations in rural areas may, in fact, be over-counted in censuses due to misclassification of rural-to-urban migrants (Table 12). Only one measure of remote or nomad census under-counts were found. The combined figure of 12% under-count was likely low based on a review of census and survey coverage of nomadic populations (Randall, 2015), though measurement of nomad under-counts are particularly difficult to estimate because nomad under-counts are not measured with PESs.

Table 12. Range of measured census under-counts in LMICs since 1990 by household type

Household type	Population type		
	Non-mobile Population (% under-count)	Mobile Population (% under-count)	Institutional (% under-count)
Urban slum			
Permanent/semi-permanent	5-59 % [(1-5)]	33-61 % [(6-8)]	Refugee / IDP: 10-15 % [(3)] Worker: Unknown
Homeless – Non-permanent	---	33-100 % [(3,6,8)]	---
Homeless – Roofless	---	64-100 % [(3,9,10)]	Homeless shelter: Unknown
Homeless – Guesthouse	---	Unknown	---
Urban non-slum			
Permanent	2-15 % [(11-20)]	---	Prison: 9 % [(3)] Military: Unknown Religious: Unknown Care home / Hospital: <1 % [(3)] Education: Unknown
Rural			
Permanent/semi-permanent	2-13 % [(11-20)]	-6 % [(8)]	Refugee / IDP: 10-15 % [(3)] Worker: Unknown Prison: Unknown Military: Unknown Religious: Unknown Education: Unknown
Remote	12 % [(21)]	---	---
Nomad	---	12 % [(21)]	---

1. (Sabry, 2010)
2. (Karanja, 2010)
3. (Carr-Hill, 2013)
4. (Carr-hill, 2017)
5. (Lucci, Bhatkal and Khan, 2018)
6. (Treiman *et al.*, 2005)
7. (Kronenfeld, 2008)
8. (Ebenstein and Zhao, 2015)
9. (Gurgel *et al.*, 2004)
10. (Stark *et al.*, 2017)
11. (Korale, 2002)
12. (UBS, 2005)
13. (Maro, 2009)
14. (NISR, 2010)
15. (BIDS, 2012)
16. (GSS, 2012)
17. (CSO, 2013)
18. (NSC, 2014)
19. (Oliveira *et al.*, 2003)
20. (SSA, 2012)
21. (Gidado *et al.*, 2013)

3.5.3 Simulating census inaccuracy in Khomas

To model inaccuracies due to census under-counts, I classified each of the simulated households in 2016, 2011, 2006, and 2001 as urban slum, urban non-slum, or rural. My household-level slum/non-slum classification was based on visual inspection of building morphology and community characteristics in Google Earth's Maxar and SPOT (40cm) imagery of every building in all four years, paying attention to features described in the household typology section (3.5.2.1). While classifying the presence and types of buildings across years, I observed that building morphologies remained virtually unchanged, with only new structures being added in and around Windhoek.

All images were mostly free of cloud cover in the study area for all years; in case of cloud cover, I used imagery collected slightly before and after the target date which was cloud free. I made the assumption that household occupancy and characteristics remained constant over time, except for a small numbers of households changing from rural to urban slum in peri-urban areas. The change from rural to urban slum type was observed in 3 simulated households between 2001 and 2006, 68 simulated households between 2006 and 2011, and 838 simulated households between 2011 and 2016.

Note that urban slum/non-slum classification of households for each year was different, though aligned with, the probability of "poor" household type assigned to urban EAs earlier in this chapter during simulation of the "true" 2016 population (Table 13). Most areas of Khomas did not change over time in terms of occupation by slum/non-slum households; however, a small number of EAs (probability weights 0.25-0.74) had few slum households in 2016, but were majority slum households in 2001 (Table 13). The manual classification of slum/non-slum households is reused later in Chapter 5 to assess household survey outcomes in urban slum versus non-slum households.

Table 13. Comparison of manually created EA-level urban "poor" probability weights in the 2016 simulated “true” population, and manually created household-level classification of “slum” households for the 2016, 2011, 2006, and 2001 census simulations

EA prob. weights for “poor” urban HH type 2016	Manual classification of slum/non-slum households for census simulations							
	2016 slum households		2011 slum households		2006 slum households		2001 slum households	
	n	%	n	%	n	%	n	%
0.00 – 0.24	60,097	0.3%	58,506	1.0%	54,669	2.6%	46,665	3.4%
0.25 – 0.74	2,838	8.0%	2,777	22.9%	2,317	75.2%	1,842	82.4%
0.75 – 1.00	34,732	99.7%	28,155	97.2%	14,920	99.4%	10,073	99.5%

Using ranges of actual census under-counts from the literature review (section 3.5.2.2), I simulated three levels of census inaccuracy – low, middle, and high – across rural, urban slum, and urban non-slum households (Table 14). I classified low inaccuracy as 2% missing among rural and urban non-slum households, and 10% missing among urban slum households. Middle inaccuracy was considered to be 5% missing among rural and urban non-slum households, and 30% missing among urban slum households. Finally, high inaccuracy was classified as 10% missing among rural and urban non-slum households, and 60% missing among urban slum households. I applied the inaccuracy rates to all households within household type such that there was no spatial pattern inherent to the simulated household under-counts.

Table 14. Realistic levels of LMIC census inaccuracy applied to the 2016, 2011, 2006, and 2001 simulated populations to simulate census under-counts (missingness)

Household type	Low inaccuracy	Middle inaccuracy	High inaccuracy
Urban slum	10%	30%	60%
Urban non-slum	2%	5%	10%
Rural	2%	5%	10%

This exercise resulted in one “true” and 15 simulated outdated-inaccurate populations to be used for analysis in the following chapters. The characteristics of these datasets are summarized in Table 15 with total number of urban slum, urban non-slum, and rural residents, reflecting the varying degrees of outdatedness (five, ten, 15 years old) and inaccuracy (low, middle, high under-count) identified in actual LMIC censuses in the literature review. In the next chapter, I use these simulated censuses to generate realistic gridded population datasets that reflect the types of estimates currently available across LMICs, and evaluate the accuracy of each gridded population dataset against the "true" reference population.

Table 15. Number of households simulated in the "true" population and 15 realistic scenarios of census outdatedness-inaccuracy, by year and household type

Year	No inaccuracy	Low inaccuracy	Middle inaccuracy	High inaccuracy
2016 (current)				
Urban slum	35,001	31,500	24,500	14,000
Urban non-slum	57,843	56,677	54,942	52,073
Rural	4,823	4,735	4,590	4,326
2011 (5 years old)				
Urban slum	28,583	25,724	20,008	11,433
Urban non-slum	55,680	54,566	52,895	50,122
Rural	5,175	5,071	4,917	4,647
2006 (10 years old)				
Urban slum	18,018	16,216	12,612	7,207
Urban non-slum	49,742	48,747	47,258	44,769
Rural	4,146	4,063	3,935	3,730
2001 (15 years old)				
Urban slum	13,149	11,834	9,204	5,259
Urban non-slum	41,700	40,866	39,612	37,514
Rural	3,731	3,656	3,547	3,373

Chapter 4: Frames

4.1 Overview

Selection of a representative household survey sample depends on having a complete sample frame of the current population. While census data may seem like an ideal sample frame, some populations may be missing, and other populations may be under-represented in LMIC censuses. As detail in Chapter 3, the most vulnerable populations, including urban poor and mobile populations, are the most likely to be under-counted and omitted in a LMIC census, and consequently, the most likely to be under-sampled in household surveys (Ezeh *et al.*, 2017; Lilford *et al.*, 2017).

Gridded population datasets are viewed as an alternative sample frame for household surveys when census data are outdated, inaccurate, or too coarsely aggregated. However, the majority of available gridded population datasets for LMICs are derived from the very outdated-inaccurate census data that survey practitioners wish to avoid. The use of a gridded population dataset as a sample frame is often done as a last resort with the approach that it is better than nothing (Galway *et al.*, 2012; Thomson *et al.*, 2012). While gridded population estimates will reflect outdated, inaccurate census counts, a gridded population model may improve the distribution of the population and provide omitted populations with some probability of sample selection; for example, in areas where new housing developments were added after the census, or if unregistered slums were omitted from the census.

This chapter evaluates the accuracy of gridded population sample frames derived from census data. I input 16 simulated EA-level and constituency-level census datasets into a Random Forest model, producing multiple 100 metre by 100 metre population estimates across Khomas. In section 4.4, I compare the accuracy of these gridded population datasets derived from outdated-inaccurate censuses against the “true” population. I also consider whether the accuracy of gridded population data can be improved through aggregation of grid cells. Section 4.5 reports the results of these analyses, and provides guidance to optimise accuracy of gridded population data derived from Random Forest models via aggregation of grid cells.

4.1.1 Research questions

I hypothesize that accuracy of gridded population datasets can be improved by aggregating small grid cells into larger grid cells. This is because, depending on the detail, accuracy, or completeness of the spatial covariates used in the Random Forest model, populations might be allocated to the correct vicinity, but not to the exact location, and thus combining neighbouring cells may smooth out local spatial errors. By the end of the chapter, I will have answered the following questions.

Research Question A: How accurate are gridded population estimates derived from outdated, inaccurate censuses in 100 metre by 100 metre grid cells as measured with RMSE and MAE?

Research Question B: How much do 100 metre by 100 metre gridded population datasets derived from outdated, inaccurate censuses data need to be aggregated in order to achieve the true population distribution (+/-20% percent) or count (+/- 10 people per hectare)?

4.2 Gridded population sample frames

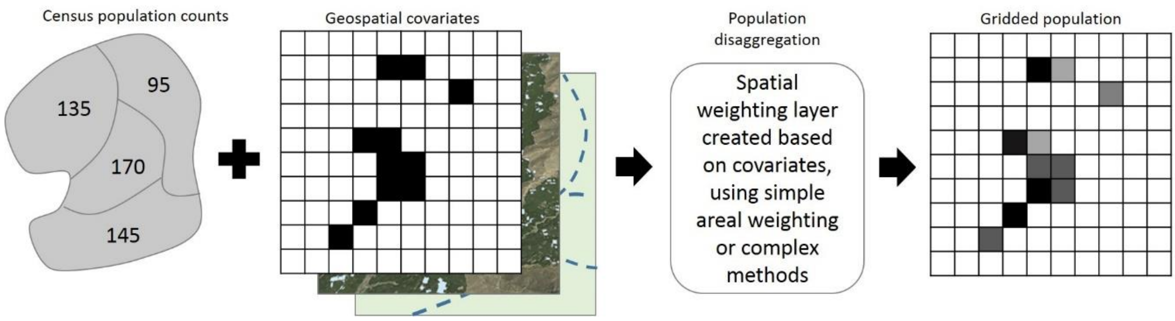
Given the prevalence of inaccuracies in LMIC census sample frames, we now turn to gridded population sample frames. This section describes top-down and bottom-up gridded population datasets derived from geo-statistical models, which may serve as alternative survey sample frames. I only consider cross-sectional gridded population datasets; gridded population datasets showing population flows derived from mobile phone data, social media, or other novel sources are not considered because sampling from dynamic population counts would require different survey methods and statistics than are typically used in household surveys (Singh and Mangat, 1996; Randall, 2015). Gridded population datasets in HICs are also excluded, including HIC “top-down” models (discussed below) (e.g., (Martin, Cockings and Leung, 2010; Martin, Lloyd and Shuttleworth, 2011)) and HIC datasets derived from aggregated administrative datasets (e.g., (European Commission, 2011, 2012)).

4.2.1 Top-down (census-based) gridded population sample frames

Top-down gridded population datasets are generated by models that disaggregate census (or other administrative) population counts from administrative areas to smaller grid cells (Figure 20). The simplest models assume a uniform distribution of population within administrative areas, while the most advanced models use spatial data to inform the spatial disaggregation. Some of these advanced models aim to reflect the average night-time residential population distribution (e.g., WorldPop-Random Forest and WorldPop-Global), while other models aim to reflect a daily “ambient” population (e.g., LandScan-Global), which can be conceived as the average between night-time residential and daytime commuter populations. To estimate population numbers beyond the last census year, population birth, migration, and death rates are used to project new population totals by administrative region (Long and McMillen, 1987).

Gridded population datasets are published as total population estimates per grid cell, where cells are generally measured in decimal degrees and are thus slightly smaller and less square-shaped toward the Earth's poles compared to the equator. I describe grid cell size in terms of metres or kilometres to facilitate understanding for a general scientific audience, though bear in mind that these are area approximations at the equator. Furthermore, when I refer to hectares, I mean grid cells that are approximately 100 metres by 100 metres at the equator. Within countries, differences in cell size are generally negligible; exceptions include Brazil and Russia with large north-south coverage. Multi-country gridded population datasets for LMICs are listed in Table 16 along with their geographic coverage, resolution, time coverage, and modelling approach. Gridded population datasets that do not have estimates after the year 2000 are excluded.

Figure 20. General workflow to create top-down, census-based gridded population data



4.2.1.1 Terms

This section introduces several technical terms *in italics* associated with gridded population data modelling. First, *dasymetric mapping* refers to the partitioning of an aerial unit into smaller units to introduce greater resolution than originally captured. Dasymetric partitions can directly reflect the ancillary data sources, for example, partitioning along land cover type boundaries. However, in the case of gridded population datasets, partitions are made along predefined grid cell boundaries, and each grid cell is assigned values from one or more ancillary data sources.

Pycnophylactic describes any disaggregation technique which ensures that the disaggregated values sum to the aggregated total; this is also called "volume preserving." All top-down gridded population datasets are pycnophylactic so that population totals in grid cells sum to the population total of the original input census data unit. This means that top-down gridded population datasets inherit outdated or inaccurate population totals at the scale of the input census data.

Table 16. Summary and comparison of top-down census-based gridded population datasets

Name	Coverage	Resolution	Years	Method	Pros	Cons
GPW (1)	Global	~1 km^2	2000-2020	Direct disaggregation *(uniform distribution)*	Easy, no ancillary spatial data needed Maintain fidelity of input data	Accuracy and precision of a given pixel is directly related to size of input census areal unit
GRUMP (2)	Global	~1 km^2	1990-2000	Informed disaggregation *(basic dasymetric such as with areal weighting using auxiliary variables)*	Maintain fidelity of input data Estimates in a given pixel are more accurate than simple disaggregation (5-7)	Few ancillary variables do not well reflect human activity across multiple physical and social environments
GHS-POP (3)	Global	~250 m^2	1975-2015			
HRSL (4)	18 countries	~30 m^2	2015			
LandScan (8)	Global	~1 km^2	2000-2017	Complex models *(intelligent dasymetric such as with Random Forest model)*	Maintain fidelity of input data Estimates in a given pixel are more accurate than both simple disaggregation or informed disaggregation (12,13)	Can feel like a “black box” because population densities do not relate directly to predictor variable
Demobase (9,10)	3 countries	~100 m^2	2002-2010			
WorldPop-LC (11-13)	57 countries	~100 m^2	2010-2015			
WorldPop-RF (14)	69 countries	~100 m^2	2010-2020			
WorldPop-Global (15)	Global	~100 m^2				

1. (Doxsey-Whitfield *et al.*, 2015)
2. (Balk *et al.*, 2005)
3. (Pesaresi, Ehrlich, *et al.*, 2016)
4. (Facebook Connectivity Lab and CIESIN, 2016)
5. (Tatem *et al.*, 2007)
6. (Linard *et al.*, 2010)
7. (Hay *et al.*, 2005)
8. (Dobson *et al.*, 2000)
9. (Azar *et al.*, 2010)
10. (Azar *et al.*, 2013)
11. (Tatem *et al.*, 2007)
12. (Linard *et al.*, 2012)
13. (Gaughan *et al.*, 2013)
14. (Stevens *et al.*, 2015)
15. (WorldPop, 2019)

4.2.1.2 Basic dasymetric methods

The earliest gridded population datasets were based on direct disaggregation approaches including simple areal weighting where each cell was assigned the same population count within a census administrative unit (e.g., GPW: (Doxsey-Whitfield *et al.*, 2015)). More accurate basic dasymetric methods soon followed using one or two ancillary datasets, such as urban settlement area boundaries or land cover type, to inform the location and density of the disaggregated population (e.g., GRUMP (Balk *et al.*, 2005), GHS-POP (Pesaresi, Ehrlich, *et al.*, 2016), HRSL (Facebook Connectivity Lab and CIESIN, 2016)).

For gridded population datasets that rely on areal weighting or basic dasymetric approaches, model errors are difficult to estimate, and to even conceptualize, as these datasets are essentially gridded representations of the input census data (Doxsey-Whitfield *et al.*, 2015). Areal weighting and basic dasymetric methods remain in use today, in part, due to the large computing power required to perform more advanced population disaggregation techniques, which become particularly demanding at a regional or global scale. Simpler modelling techniques and limited covariates are also used to isolate and study relationships between population density and other factors. However, when visually compared to actual population counts, areal weighting and basic dasymetric methods consistently produce less accurate cell-level population estimates than advanced dasymetric modelling techniques, which exploit multiple ancillary variables related to human activity (Azar *et al.*, 2013; Stevens *et al.*, 2015).

4.2.1.3 Sources of error

In addition to the modelling algorithm itself, errors in gridded population estimates are introduced by: (i) inaccuracy of the input population data, (ii) the geographic scale of the input population data (e.g., census tract versus district), (iii) the age, accuracy, completeness, and type of ancillary data, (iv) the nature of the relationship between ancillary data and population density, and (v) the geographic scale of the output grid (e.g., estimates for 1 kilometre by 1 kilometre cells will almost always be more accurate than estimates for 100 metre by 100 metre cells). Of these, the strongest predictor of accuracy in top-down gridded population models is the resolution of the input population data, with population age also playing a big role (Hay *et al.*, 2005).

Generally, the input population data are third, fourth, or fifth-level administrative units from the most recent census. The finest scale administrative data available in LMICs are georeferenced household locations or census EAs, though these are rarely available. The differences in accuracy between basic dasymetric methods and advanced dasymetric modelling techniques diminish with more detailed input population data (Hay *et al.*, 2005). Some LMICs do not have digitized census EAs, and survey fieldwork in these contexts involves asking local leaders from communities where EA boundaries fall, which can result in over- and under-counting near EA boundaries. Furthermore, differences in administrative boundaries used to collect and report the data versus boundaries used in the modelling process can lead to population misallocation. The lack of detail in population data is the motivation for developing gridded population models to begin with, and thus I focus on more accurate advanced dasymetric methods: WorldPop-Land Cover, WorldPop-Random Forest, WorldPop-Global, LandScan-Global, and Demobase.

4.2.1.4 Advanced dasymetric methods

WorldPop-Land Cover (hereafter called WorldPop-LC) has been supplanted by WorldPop-Random Forest and WorldPop-Global datasets, however much of the initial evaluation of gridded population datasets was performed on WorldPop-LC. WorldPop-LC datasets were produced for 100 metre by 100 metre cells in 57 African, Asian, and Latin American countries for 2010 and 2015. This modelling technique used boundaries of settled areas and several land cover type datasets to develop detailed land cover classes (Linard, Gilbert and Tatem, 2011; Gaughan *et al.*, 2013). A subset of the census input population data were used to calculate the average population density in each land cover class, then land cover class population densities were used as weights to distribute census population totals across the whole country. Finally, accuracy was assessed at the scale of the input census data by comparing the modelled population counts with the reserved non-subset census counts (Linard, Gilbert and Tatem, 2011).

WorldPop-Random Forest (hereafter called WorldPop-RF), supplanted by WorldPop-Global, produced 100 metre by 100 metre population estimates by country in five-year periods between 2000 and 2015 for 69 countries in Africa, Asia, and Latin America. In WorldPop-RF models, input census data and all available covariate data were aggregated at two scales to test and tailor the model to local areas, producing a population probability weight for each grid cell which was then used to dasymetrically disaggregate census population counts (Stevens *et al.*, 2015). The WorldPop-RF datasets superseded the earlier WorldPop-LC datasets because the Random Forest model was found to be more accurate than land cover weights (Stevens *et al.*, 2015). WorldPop-RF methods are well documented in the public domain, and include source code to recreate datasets (Stevens *et al.*, 2015).

WorldPop-Global is a free, open-source 100 metre by 100 metre dataset of population for individual countries covering the entire global based on the same modelling techniques as WorldPop-RF. WorldPop-Global supplanted WorldPop-RF because it includes annual population estimates between 2000 and 2020 for all countries, with harmonized gridded boundaries across the globe (Lloyd *et al.*, 2019). In a few countries, WorldPop-Global has fewer covariate datasets than the WorldPop-RF models which, in theory, would produce slightly less accurate results if the models were run for the same country at the same time. However, WorldPop-Global has been updated much more recently (2019) than WorldPop-RF models (2013-2016), so the accuracy and completeness of dynamic covariate datasets, such as roads and points of interest from OpenStreetMap, are more accurate in the WorldPop-Global datasets (WorldPop, 2019).

LandScan-Global is a 1 kilometre by 1 kilometre daily ambient population estimate for the globe from 2000 to the current year. The probability weights matrix used for dasymetric disaggregation is generated with co-kriging, a multivariate spatial interpolation method, and four ancillary datasets with global coverage: roads, slope, land cover, and night-time lights (Dobson *et al.*, 2000). To account for economic, physical, and cultural differences that might affect the relationship between covariate datasets and population density in local areas, a LandScan analyst assigns relative weights by location to each of the ancillary datasets to adjust the model. While these weights are made available, LandScan's process for generating them is opaque. LandScan-Global does not publish model errors related to the co-kriging model, nor does it include metadata for the ancillary data used. LandScan-Global is a commercial dataset and made free to US Federal Government agencies and to some humanitarian, education and commercial organizations, upon request (ORNL, 2017).

Demobase 100 metre by 100 metre datasets have been created by the US Census Bureau for Haiti, Pakistan, Rwanda, and South Sudan for 2003, 2010, 2000-2013, and 2017, respectively. The probability weights matrix used for dasymetric disaggregation is based on ancillary variables derived entirely from satellite imagery. This modelling approach is considerably more complex than basic dasymetric datasets modelled from built area types; in Demobase the population weights matrix is modelled from dozens of ancillary variables generated through semi-automated classification of high resolution satellite imagery and supervised classification of medium resolution satellite imagery (Azar *et al.*, 2013). Model errors are derived during the classification process based on pixels that were classified by Demobase analysts and withheld from the model. Demobase data are free and publicly available (USCB, 2013), and the methods are clearly documented in the public domain (Azar *et al.*, 2010, 2013).

4.2.1.5 Accuracy of advanced dasymetric methods

Accuracy of gridded population datasets derived with advanced dasymetric methods are often measured with mean absolute error (MAE), a measure of precision, and root mean square error (RMSE), a measure of error magnitude that penalises large errors. In MAE, the absolute difference between the predicted population and actual population in each census unit is calculated and averaged.

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

In RMSE, the difference in each census unit is taken and squared. Next, the squared differences are averaged and the square root is taken to calculate a mean difference across census units, which makes this measure sensitive to outliers and large differences.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

There are no studies comparing *model accuracy* of all WorldPop-LC, WorldPop-RF, WorldPop-Global, LandScan-Global, and Demobase for reasons I will describe in the next paragraph. However, there are a few studies comparing *output accuracy* from these models, though the benefits of each modelling approach are difficult to discern because both the models and the input population datasets vary in most comparisons. For example, the producers of Demobase compared population estimates in Pakistan derived from tehsils (a low-level administrative unit approximately the size of townships in the United States) to LandScan-Global population estimates, and found lower RMSE in Demobase when both datasets were aggregated to census EA boundaries and compared to 1998 Pakistan census EA counts (Azar *et al.*, 2013). Without running the LandScan model with the same tehsil population input data, it is not clear whether the greater accuracy of Demobase was due to the modelling technique, or due to greater resolution of the input population data. Similarly, WorldPop-LC compared population estimates for Somalia with LandScan-Global and found much lower RMSE in the WorldPop-LC dataset. However, this was likely because the validation data were the population data used as an input to the WorldPop-LC model (Linard *et al.*, 2010).

A key challenge of comparing these data sources is that LandScan-Global methods and datasets are not publically available. LandScan-Global's producer, Oakridge National Laboratories with support from the United States Department of Defence, does not provide sufficient information about their modelling methods and input data to enable replication (Stevens *et al.*, 2015), thus it is not possible for researchers to generate LandScan datasets for the purpose of accuracy comparison.

LandScan-Global's internal accuracy evaluations have been against earlier, much simpler methods developed by the US Government to model global gridded populations between 1965 and 1995 (Dobson *et al.*, 2000). With regard to Demobase, the high resolution imagery needed for the model covariates is not only cost-prohibitive to obtain, it is computationally demanding to process at a regional and global scale.

WorldPop-Global was released in 2019, and no studies were yet published at the time of this writing that compared accuracy of WorldPop-Global with other gridded population datasets. However, given that WorldPop-Global uses the same modelling approach as WorldPop-RF, albeit with fewer covariates, similar model accuracy to WorldPop-RF is assumed. I confirmed this assumption by comparing EA and cell-level accuracy for all top-down gridded population datasets in approximately the same year (2014/15) in Oshikoto, Namibia where I simulated a realistic population with colleagues (Thomson, Kools and Jochem, 2018). In Oshikoto, WorldPop-RF and WorldPop-Global were more accurate than other gridded population estimates, and their level of accuracy was similar (see Appendix C for details).

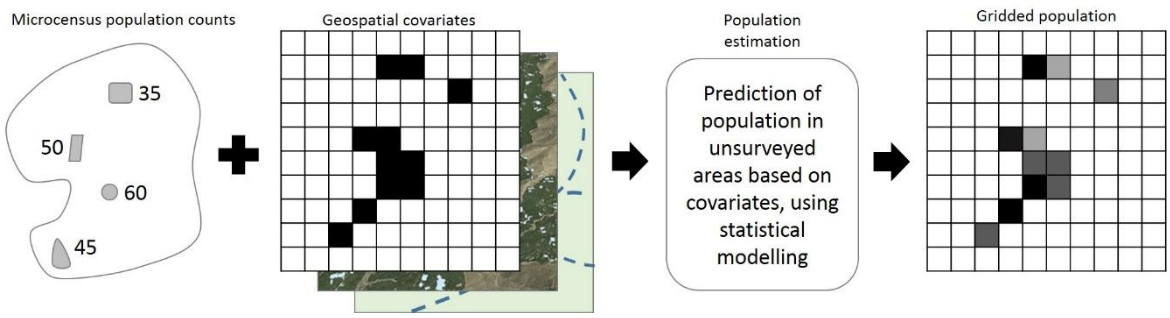
For the purposes of selecting a top-down gridded population modelling approach to evaluate the accuracy of a gridded population estimates at the cell-level, the WorldPop-RF / WorldPop-Global Random Forest technique was a natural choice for several reasons. First, WorldPop's Random Forest methods are well documented and assessed against WorldPop-LC and simpler disaggregation models such as GPW (Gaughan *et al.*, 2013; Stevens *et al.*, 2015) and GRUMP (Gaughan *et al.*, 2013; Stevens *et al.*, 2015). Second, Random Forest models allow for estimation of model errors. Third, the WorldPop Global project provides all model covariates pre-processed at the 100 metre by 100 metre scale (WorldPop, 2018a; Lloyd *et al.*, 2019). The 100 metre by 100 metre resolution of the WorldPop-RF and WorldPop-Global datasets are a major advantage over 1 kilometre by 1 kilometre datasets in terms of survey feasibility, as their finer resolution means that cells can be combined to meet various sample design requirements. Finally, WorldPop-RF and WorldPop-Global both model the residential night-time population rather than ambient population, which is appropriate for household surveys.

4.2.2 Bottom-up (micro-census-based) gridded sample frames

Researchers are experimenting with gridded population modelling techniques that are census-independent which would be especially useful where census data are unavailable, outdated, or inaccurate. Bottom-up gridded population modelling methods involve establishing a statistical relationship between population density and spatial covariates in a sample of small areas (Figure 21). This relationship is used to predict population densities across a whole country based on spatial covariates alone (Tatem, 2017; Wardrop *et al.*, 2018).

Bottom up gridded population mapping methods are still being developed, evaluated, and scaled up. Examples of bottom up maps include Sierra Leone (Hillson *et al.*, 2014), Nigeria (Weber *et al.*, 2018), and Afghanistan (Wardrop *et al.*, 2018). The GRID3 project, funded by Gates Foundation and the UK Department for International Development, and the LandScan-HD project, funded by the US Oakridge National Laboratories, are both currently working on bottom-up gridded population estimates in 100 metre by 100 metre grid cells for multiple LMICs (CIESIN *et al.*, 2018). A bottom up map in Namibia was not available at the time of this writing, and was thus not considered in this thesis.

Figure 21. General workflow to create “bottom-up,” census-independent gridded population data



4.2.3 WorldPop – Random Forest Model

In this section, I detail the Random Forest model used in the WorldPop-RF and WorldPop-Global datasets, which I use to generate gridded population datasets in this thesis. The WorldPop-RF / WorldPop-Global model generates per-pixel population weights which are applied to a census or projected census dataset to dasymetrically disaggregate population counts from administrative units to cells such that the population counts are pycnophylactic. The Random Forest model is a non-parametric ensemble machine-learning algorithm that grows a “forest” of decision trees through the modelling process (Breiman, 2001). *Machine-learning* describes a computational algorithm which “learns” patterns in very large datasets and uses the acquired information to improve predictions. Each Random Forest tree is a model of the potential relationships between the spatial covariates and the census population counts, and this relationship is not tied to linearity constraints. A non-parametric machine learning algorithm is one which is not based on an assumed probability distribution in the data, nor does it have a set number of parameters. The Random Forest model is non-parametric in the sense that the number of decisions within each tree is not fixed; rather, the number of decision points increases with the number of ancillary variables. However, users of the Random Forest model will note that it has just two modelling parameters, which I will describe.

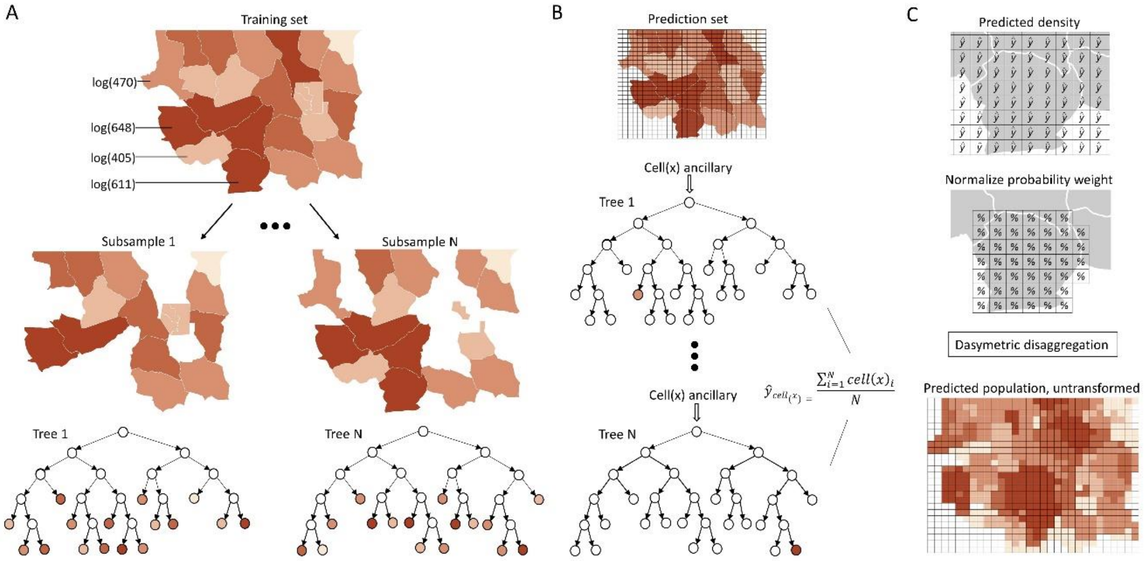
The Random Forest model is part of a family of *ensemble* machine learning algorithms. Ensemble algorithms are comprised of many sub-models, or decision trees, in this case. Boosted Regression Trees is a related ensemble machine learning algorithm (Breiman, 1996). Both Random Forest and Boosted Regression Trees build decision trees from ancillary data. Each decision point, or *node*, of a tree is associated with one ancillary variable, and each node has two branches. The node for a temperature variable might split, for example, at temperature >= 17°C or temperature < 17°C. Random Forests are essentially Boosted Regression Trees with an added layer of randomness introduced into the bagging process (Liaw and Wiener, 2002).

Bagging is the process used to create decision trees, including calculation of the split value at each node. Bagging stands for "bootstrap aggregating" and it means that a portion of observations are randomly sampled with replacement from the training dataset, and successive decision trees are independently built such that the n^{th} decision tree does not depend on earlier decision trees. In Boosted Regression Trees, the split value at each node is based on an optimal split calculation using all ancillary variables in the dataset. In Random Forest models, the optimal split values are calculated from a subset of ancillary variables randomly selected at each node. This added element of randomness prevents overfitting, avoids a variable selection process, and addresses collinearity issues (Breiman, 2001).

The two parameters in a Random Forest model are the number of variables in the random subset at each node, and the number of trees in the forest (Liaw and Wiener, 2002). The WorldPop-RF / WorldPop-Global algorithm builds decision trees using census unit population counts and spatial covariates processed at the census unit scale (Figure 22) (Stevens *et al.*, 2015). Data are initially prepared by census unit such that for each census unit identifier, the population is known and multiple spatial ancillary data values are joined. WorldPop-Global uses a standard set of 24 covariates from 10 sources (Lloyd *et al.*, 2019) in which values are converted to categorical values (e.g., land cover type, urban-rural) or continuous values (e.g., intensity of night-time lights, distance to roads) for each administrative area.

In the second step of the Random Forest model, all of the ancillary datasets are prepared in 100 metre by 100 metre cells. The second step uses the split values of each classification tree developed in the first step to parameterize a corresponding regression model to predict population density within cells (Stevens *et al.*, 2015). For each cell, the predicted population values from all regression models are averaged to make a single population estimate, though these population estimates are not pycnophylactic.

Thus WorldPop-RF / WorldPop-Global perform a third step outside of the Random Forest model to normalize cell-level predicted population densities (Stevens *et al.*, 2015). This is done by summing the predicted population densities by census unit, and calculating the proportion of census unit population located in each cell. Multiplication of this proportion by the census unit population dasymetrically disaggregates the census counts and ensures they are pycnophylactic.



A) Each decision tree in the ensemble is built upon a random bootstrap sample of the log-transformed population and ancillary data by administrative unit. **B)** Population density prediction for each cell $\hat{y}_{cell_{(x)}}$ is based on an average of the individual trees. **C)** Predicted cell densities are normalized by administrative unit and used to dasymetrically disaggregate log-transformed administrative unit population, then untransformed to predict population per cell.

Figure 22. WorldPop-RF and WorldPop-Global classification, regression, and weighting process using Random Forest models with remotely-sensed and ancillary data

Benefits of the Random Forest model are: (1) it is non-parametric so it can model complex relationships between population density and ancillary variables without *a priori* assumptions, (2) prediction errors can be estimated at the census unit scale, (3) the model accommodates both continuous and categorical variables allowing many diverse covariate datasets to be used during the modelling process, and (4) the model is capable of dealing with collinear covariates and non-linear associations.

A limitation of the Random Forest model is that predictions of the outcome are bounded by the range of observed values in the original census training data. This has two consequences for the prediction of population in WorldPop-RF / WorldPop-Global datasets (Stevens *et al.*, 2015). First, the number of values may be limited, resulting in cell estimates that lack “realistic” gradations and instead have abrupt edges within and around population centres. Second, the lower and upper range of average population density in census units may not represent the lower and upper range of population density in cells. This limitation is particularly relevant in countries with coarse input census data. In these circumstances, a combined model for two countries can be built – the country of interest with course census data, and a second country in the same region which has more detailed census data – to introduce a wider, and more realistic, range of population values which can be assigned to 100 metre cells (Gaughan *et al.*, 2015).

Another limitation of the WorldPop-RF / WorldPop-Global model results from the choice to transform population counts to a log-scale (Stevens *et al.*, 2015). As a result of the log transformation, these datasets have a non-zero population prediction for all cells, and thus a small fraction of the census population is allocated to uninhabited areas, for example in deserts or forests, where the estimated cell population might be 0.00001 persons.

Finally, many of the covariate datasets used to model population counts in WorldPop-RF / WorldPop-Global 100 by 100 metre cells are only available at a coarser resolution, limiting cell-level accuracy in the model. Land cover classes are available at approximately 300 by 300 metre resolution (ESA-CCI, 2017a); and night-time lights (NOAA, 2012, 2017; Zhang, Pandey and Seto, 2016), distance to nature reserves (UNEP-WCMS and IUCN, 2016), travel time to cities (Weiss *et al.*, 2018), and precipitation and temperature (Fick and Hijmans, 2017) are all published at approximately 900 by 900 metre resolution. The only covariates available at a similar resolution to the output grid cells are distance to major roads (OpenStreetMap contributors, 2000), intersections (OpenStreetMap contributors, 2000), waterways (OpenStreetMap contributors, 2000), inland water bodies (ESA-CCI, 2017b), coastlines (CIESIN, 2016), urban areas (Pesaresi, *et al.*, 2016), and built settlements (DLR Earth Observation Center, 2017), as well as slope and elevation (de Ferranti, 2017a, 2017b). None of the finer resolution datasets are particularly informative about the distribution of population within urban areas; and those datasets which might be informative (e.g., distance to roads or intersections) are likely to perform poorly in slums either because roads are absent or not well mapped in OpenStreetMap. Coarse datasets which are likely to correlate with population density and distribution, such as night-time lights, will result in a “halo” effect with population being disaggregated to cells in the vicinity of a true high dense cell, but not necessarily the correct cell itself.

4.3 Simulated gridded population sample frames

In Chapter 3, I simulated 16 realistic census scenarios from household points of a “true” population in Khomas, Namibia. The scenarios represent varying degrees of outdatedness (zero, five, ten, and 15 years) and inaccuracy (low, middle, and high levels of missingness among urban slum, urban non-slum, and rural populations). In this chapter, I aggregate each of the simulated household populations to EA and constituency (second-level administrative unit) boundaries, and simulate a total of 32 realistic gridded population datasets. It is important to evaluate the accuracy of gridded population datasets derived from different census aggregations because cell-level accuracy of all top-down gridded population datasets is highly influenced by aggregation of the input population. The names of all 32 simulated gridded population input datasets are listed in Table 17.

Table 17. Names of all simulated census scenario datasets, reflecting different years of outdatedness, realistic levels of population under-counts (inaccuracy), and levels of model input aggregation (EA versus constituency)

Year	No inaccuracy	Low inaccuracy	Middle inaccuracy	High inaccuracy
2016 (current)	EA_2016_true Const_2016_true	EA_2016_L Const_2016_L	EA_2016_M Const_2016_M	EA_2016_H Const_2016_H
2011 (5 years old)	EA_2011_true Const_2011_true	EA_2011_L Const_2011_L	EA_2011_M Const_2011_M	EA_2011_H Const_2011_H
2006 (10 years old)	EA_2006_true Const_2006_true	EA_2006_L Const_2006_L	EA_2006_M Const_2006_M	EA_2006_H Const_2006_H
2001 (15 years old)	EA_2001_true Const_2001_true	EA_2001_L Const_2001_L	EA_2001_M Const_2001_M	EA_2001_H Const_2001_H

4.3.1 Methods

I used the same methods and parameters as WorldPop-RF / WorldPop-Global to prepare covariates and population data, run Random Forest models, and to reweight population densities after modelling. The steps are detailed below and visualized in Figure 23 and Figure 24.

Step 1: Population in Khomas. To prepare population counts in Khomas for Random Forest modelling, I aggregated the number of household members to EA and constituency boundaries for each of the 16 simulation scenarios, resulting in 32 shapefiles with scenario-specific population totals. As mentioned in section 3.4.1.4, I combined two EAs in Windhoek city centre with a neighbouring EA which had similar housing characteristics in satellite imagery because zonal statistics could not be calculated for their small narrow shapes; this reduced the number of EAs in Khomas from 922 to 920.

Step 2: Population in Namibia without Khomas. The number of constituencies (n=10) inside Khomas did not provide a sufficient number of observations to run a Random Forest model on their own. To boost the number of observations in the Random Forest training dataset, I tested inclusion of another population dataset for the rest of Namibia on model results. Population totals outside of Khomas represented actual 2011 census EA and constituency population counts, projected to 2016. These numbers were derived by aggregating WorldPop-Global 2016 population estimates based on 2011 Census EA populations (WorldPop, 2019). The additional data comprised of 4,575 EAs, and 97 constituencies outside of Khomas, and were generated in ArcGIS 10.5 using zonal statistics (ESRI, 2018) (Figure 23).

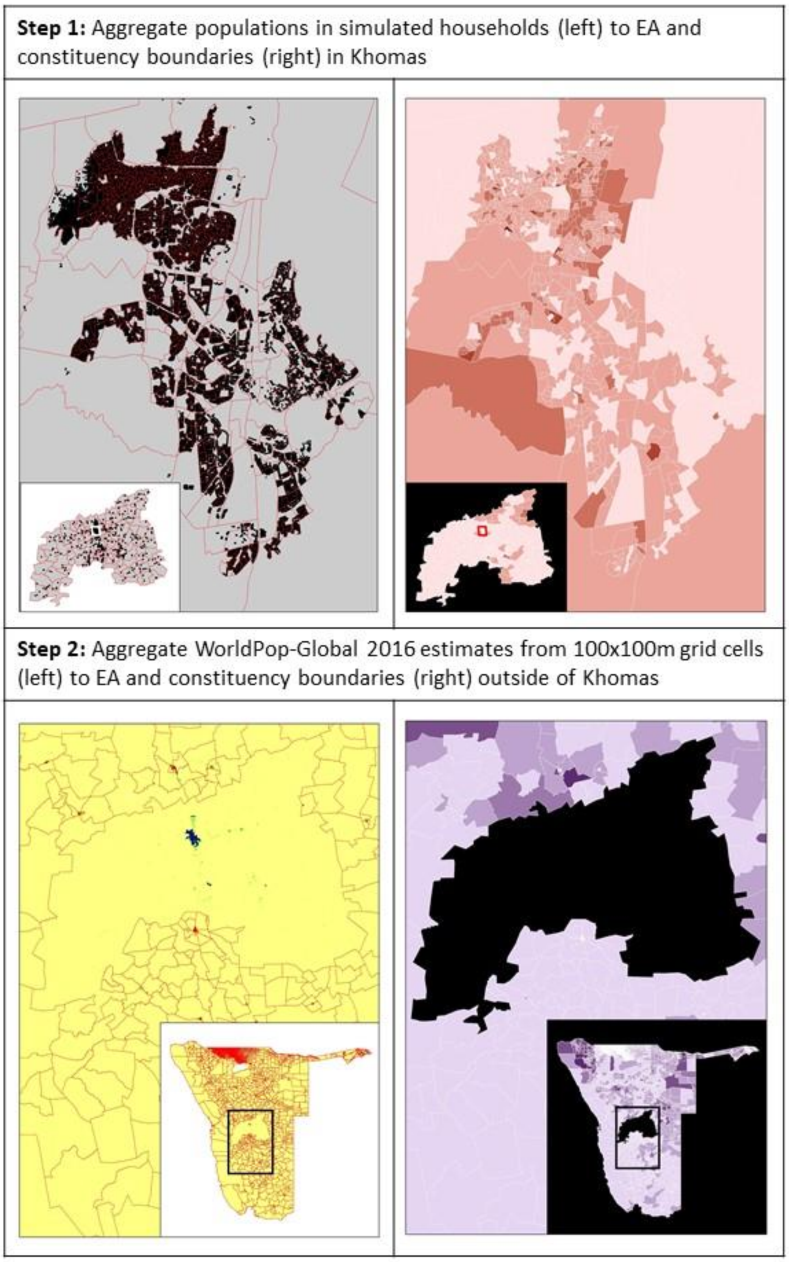


Figure 23. First two steps in WorldPop-RF/WorldPop-Global gridded population modelling workflow: Preparing population data for Random Forest models

Step 3: Combined Namibia population. Using the ArcGIS 10.5 merge tool, I combined 2016 EA and constituency population estimates outside of Khomas with each of the 32 simulated EA and constituency population datasets, resulting in 5,495 EA estimates, and 107 constituency estimates (Figure 24). I was not concerned that population totals outside of Khomas were all for 2016, and that they did not include population missingness per my simulated population scenarios. The sole purpose of including these counts was to provide the Random Forest model with a sufficient number of observations and sensible population densities to train the model. I only interpreted and analysed model outputs for Khomas. This step is in-line with the practice of combining neighbouring area population data to train Random Forest models when the target area population data is coarse and thus has few units (Stevens *et al.*, 2015).

Step 4: Covariate zonal statistics. Nearly all of the covariates that I used to train Random Forest models were pre-processed by the WorldPop-Global project and downloaded from the WorldPop FTP server (WorldPop, 2018a). In total, I used the same 24 covariates as WorldPop-Global including distance to road, road intersection, coastline, land cover type, built area, and protected area, as well as topography (elevation), slope, and night-time light intensity (Table 18). The WorldPop-Global covariates are resampled to a common 3-arc second (approximately 100 metre by 100 metre) raster and projected to GCS WGS 1984 (Lloyd *et al.*, 2019). Due to license redistribution restrictions, average annual temperature and average annual precipitation were not available via the WorldPop FTP server. To generate these covariates, I used the EA-level zonal statistics calculated by WorldPop-Global (WorldPop, 2018b), and then rasterized those datasets to the same dimension as other covariates in R 3.5.2 using the `rasterize` function. Finally, I calculated mean zonal statistics for each covariate by EAs (which included my boundary modifications) and constituencies for the whole of Namibia in R 3.5.2 using the `zonal` function.

Note that some of the underlying datasets used in this Random Forest model of gridded populations are the same datasets used in Chapter 3 to simulate the “true” population; however, the datasets are processed differently in the two analyses resulting in different covariates. In Chapter 3’s simulation, covariates represented minimum, maximum, and mean values of covariates in 100 metre pixels summarised to a two kilometre moving window. In the gridded population models in this chapter, covariates are averaged by EA and later resampled to 100 metre by 100 metre grid cells. Other important differences between the analyses are (1) the Random Forest model in the Chapter 3 simulated population was used to classify household types (but their density and distribution was predetermined by manual digitisation of building point locations), whereas the Random Forest model in this chapter is used to estimate population density and distribution; and (2) the Random Forest household type probability surfaces were reweighted by probabilities (proportion) of poor and non-poor households in each EA, adding noise to the results. If use of the same underlying datasets induced any correlation between household types (Chapter 3) and population density and distribution (Chapter 4), the correlations likely reflected actual phenomena, and not statistical artefacts (Engstrom, 2016).

Step 5: Random Forest model and reweighting. Random Forest models were produced in R 3.5.2 using the `randomForest` function (Liaw and Wiener, 2002) on the University of Southampton Iridis 5 high performance computing cluster. Before running the models, the area of each EA and constituency was calculated in hectares, and average population densities were calculated for each EA and constituency across the 32 population datasets by dividing total population by area.

These raw population densities were highly skewed to the right and many EAs in protected areas and deserts had zero population. To provide the Random Forest model with an even distribution of population densities to ensure the algorithm found optimal splits in the data, EAs with zero population were dropped. Next, a log function was applied to population density in each EA and constituency census scenario to transform them into normal distributions. During development of the WorldPop-RF / WorldPop-Global modelling approach, other approaches and transformations were tested including square root and $\log_{10}$, however, this combination of dropping zero values and log transforming densities consistently resulted in the best fitting models across settings (Stevens *et al.*, 2015).

To evaluate the effect of including areas outside of Khomas on the model results, and to test whether the default 500 trees was sufficient, I ran three versions of the Random Forest Model on EA-level datasets: (1) Khomas EAs with 500 trees (n=920), (2) Khomas EAs with 1000 trees (n=920), and (3) EAs for Khomas combined with rest of Namibia with 500 trees (n=4,575). After confirming that 500 trees were sufficient, and that the inclusion of areas outside of Khomas had minimal effect on the Random Forest model results (detailed in next section), I used the third model results based on the combined population in Khomas and the rest of Namibia. I additionally ran a model that combined simulated constituencies in Khomas with 2016 constituency populations for the rest of Namibia.

Each model involved the following steps. First, an initial tuning model was run using the non-zero log population counts and mean zonal statistics of all 24 covariate datasets. The models were parameterized with the same values as WorldPop-RF and WorldPop-Global including number of variables in the random subset at each node (*mtryStart* = n variables ÷ 3) and number of trees in the forest (*ntreeTry* = n observations ÷ 20) (Stevens *et al.*, 2015). By using the `tuneRF` function, I identified which covariates contributed positive increases to node purity (e.g., increase of variance explained), and the optimal number of covariates to randomly select at each branch of the model. Each model was then retuned using the reduced set of covariates which increased node purity (Figure 24).

After the second model tuning, the remaining covariates and corresponding trees were used to predict population estimates in 100 metre by 100 metre grid cell based on the original rasterized covariate values (Figure 24). Note that temperature and precipitation covariates in this step may have lacked spatial detail compared to other covariates, as they represented EA-level means.

In the models based on Khomas data only, cell estimates were reweighted so that cell-level estimates summed to the original EA population totals used to fit the models. In the model that combined Khomas with the rest of Namibia, areas outside of Khomas were first dropped before

Khomas cell estimates were reweighted, so that cell-level estimates summed to the original EA or constituency population totals used to fit the model in Khomas only:

$$cell(pop)_{new} = \frac{cell(pop)_{old} \times unit(pop)_{sim}}{\sum_{unit} cell(pop)_{old}}$$

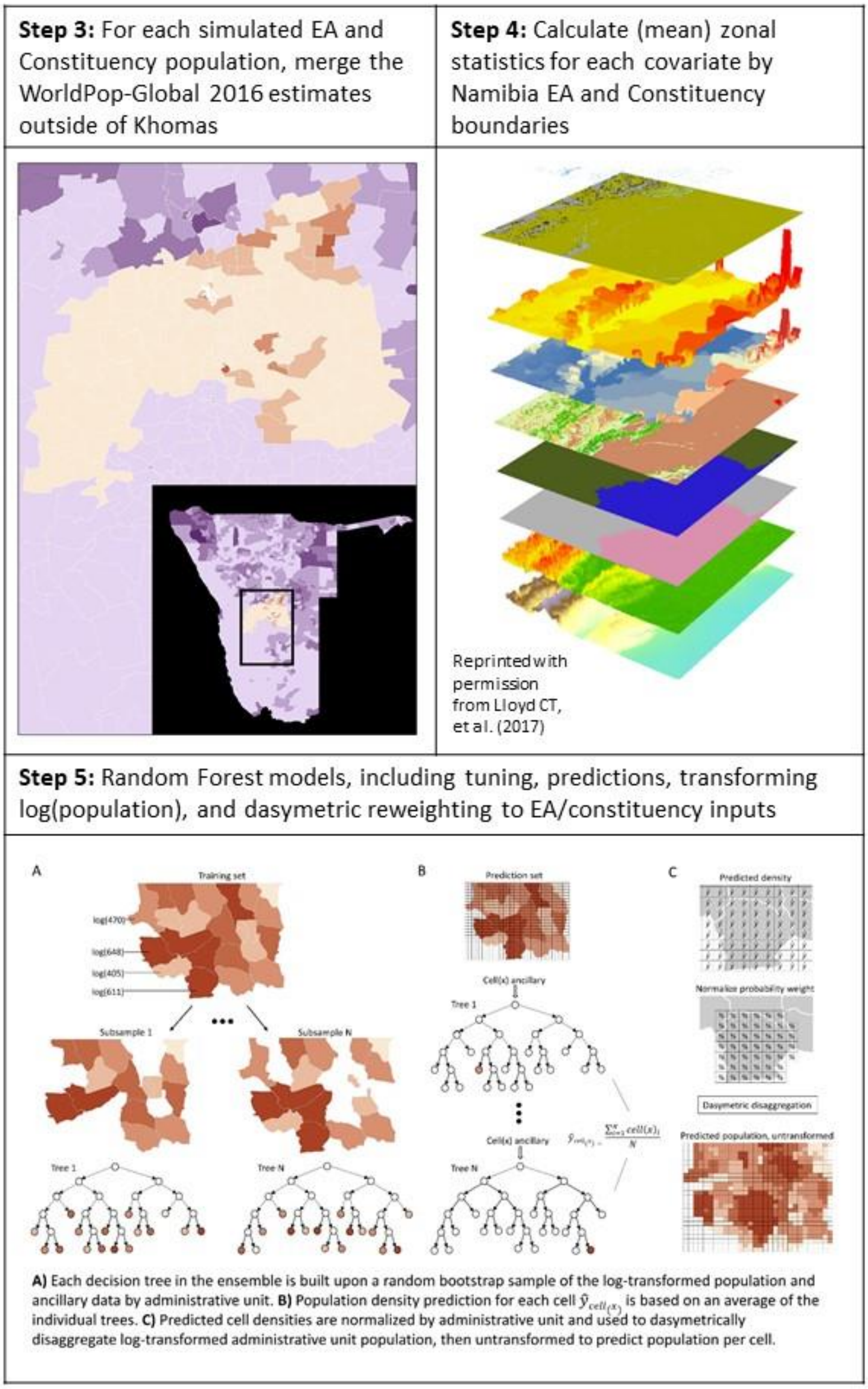


Figure 24. Last three steps in WorldPop-RF/WorldPop-Global gridded population modelling workflow: Preparing covariates and running Random Forest models

Table 18. Covariate data sources for Random Forest gridded population estimates

Name	**Description (Year)**	**Original scale**	**Original source**
cov_road	Distance to OSM major roads (2016)	Vector, <30 m	OpenStreetMap [1]
cov_intsec	Distance to OSM major road intersections (2016)	Vector, <30 m	OpenStreetMap [1]
cov_waterw	Distance to OSM major waterways (2016)	Vector, <30 m	OpenStreetMap [1]
cov_wdpa	Distance to IUCN nature reserve (2000-17)	30” (~900 m)	UNEP-WCMS & IUCN [2]
cov_viirs	Resampled VIIRS night-time lights (2012-2016)	30” (~900 m)	NOAA [3]
cov_dmsp	Resampled DMSP-OLS night-time lights (2011)	30” (~900 m)	NOAA & Zhang, et al. [4,5]
cov_tt50k	Resampled travel time to cities of 50,000+ (2000)	30” (~900 m)	Weiss, et al. [6]
cov_001	Distance to cultivated areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_040	Distance to woody areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_130	Distance to cultivated areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_140	Distance to herbaceous areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_150	Distance to sparse vegetation areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_160	Distance to aquatic vegetation areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_190	Distance to urban areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_200	Distance to bare areas (2015)	9” (~300 m)	ESA CCI – LC [7]
cov_cciwat	Distance to ESA-CCI-LC inland waterbodies (2000-12)	4.5” (~150 m)	ESA CCI [8]
cov_slope	SRTM-based slope (2000)	3” (~90 m)	de Ferranti [9,10]
cov_topo	SRTM-based elevation (2000)	3” (~90 m)	de Ferranti [9,10]
cov_coast	Distance to open-water coastline (2000-20)	3” (~90 m)	CIESIN [11]
cov_ghsl	Distance to urban area (2012)	1.26” (~38 m)	Pesaresi, et al. [12]
cov_guf	Distance to settlement built-up areas (2012)	2.8” (~84 m)	DLR EOC [13]
cov_bsgme	Distance to built settlement expansion (2016)	3” (~90 m)	Nieves, et al. [14]
cov_prec	Average total annual precipitation (1970-2000)	30” (~900 m)	Fick and Hijmans [15]
cov_temp	Average annual temperature (1970-2000)	30” (~900 m)	Fick and Hijmans [15]

1. (OpenStreetMap contributors, 2000)
2. (UNEP-WCMS and IUCN, 2016)
3. (NOAA, 2012)
4. (NOAA, 2017)
5. (Zhang, Pandey and Seto, 2016)
6. (Weiss *et al.*, 2018)
7. (ESA-CCI, 2017a)
8. (ESA-CCI, 2017b)
9. (de Ferranti, 2017b)
10. (de Ferranti, 2017a)
11. (CIESIN, 2018)
12. (European Commission, 2017)
13. (DLR Earth Observation Center, 2017)
14. (Nieves *et al.*, 2020)
15. (Fick and Hijmans, 2017)

Datasets accessed via the WorldPop FTP server (WorldPop, 2018a). Data processed by the WorldPop team at 3 minute (~100x100m resolution) and projected to GCS WGS 1984 (Lloyd, Sorichetta and Tatem, 2017; Lloyd *et al.*, 2019).

4.3.2 Results

The comparison of 500 and 1000 trees in the Khomas EA-level Random Forest models showed no effect of increasing the number of trees, thus I decided to use 500 trees in all models (Table 19).

The comparison of models run with Khomas simulated EA-level populations only, and Khomas EA-level simulated populations combined with 2016 estimates for the rest of Namibia, also resulted in minimal differences. Due to the greater number of observations, the models that included data for all of Namibia performed better in terms of mean square error (MSE) and coefficient of variance (CV) (square root of MSE divided by number of households), and explained roughly 1% more of the overall variance in population densities (Table 19). Given the minimal effect on MSE and percent variance explained of including population outside of Khomas to train the EA-level models, I chose to use 2016 population in constituencies outside of Khomas to train the constituency-level models for Khomas's 10 constituencies. To maintain comparability between EA- and constituency-level results throughout the rest of this thesis, I used the EA-level model outputs trained with data for all of Namibia. Again, only the gridded population estimates inside Khomas were analysed.

The Random Forest models with EA-level input population data performed slightly better (approximately 95% variance explained) (Table 19) than models with constituency-level input population data (approximately 94% variance explained) (Table 20), which is to be expected due to the finer resolution of EAs. The `randomForest` function provides MSE results for each model, enabling assessment of model error at the scale of the input population (Liaw and Wiener, 2002). MSE in the EA-level models (0.47 to 0.50) (Table 19) were higher than constituency-level models (0.34 to 0.38) (Table 20) because population totals at the EA-level vary more than at the constituency-level.

Across models, CV increased as the input population was increasingly outdated and as the level of population missingness increased, as was expected due to mismatches between population density and spatial covariates (Table 19, Table 20). Table 21 summarises covariate importance scores (see section 3.4.1.4) for population estimates derived from select EA- and constituency-level models. See Appendix D for plots of covariate importance and observed versus predicted population in each of the 32 models that included data outside of Khomas for training, as well as for the 16 EA-level models (500 trees) that were trained on data from Khomas only.

Before advancing to sample frame accuracy assessments, I confirmed that the rescaled gridded population estimates aggregated to the correct simulated census totals in EAs or constituencies.

Table 19. Comparison of Random Forest model fit statistics for EA-level input census scenarios

*Bold indicates model outputs used in further analyses, * multiplied by 10,000,000*

Scenario model	Coverage	Number of trees	Number of splits	Mean Square Error (MSE)	Coefficient of variation (CV)*	Percent variance explained
EA_2016_true	Khomas	1000	6	0.6452042	82.24	94.85
		500	6	0.6341473	81.54	94.94
	Namibia	**500**	**5**	**0.4754283**	**3.57**	**95.84**
EA_2016_L	Khomas	1000	5	0.6390794	86.04	94.86
		500	6	0.6465398	86.54	94.80
	Namibia	**500**	**8**	**0.4763202**	**3.58**	**95.81**
EA_2016_M	Khomas	1000	6	0.6197341	93.68	94.92
		500	6	0.6212137	93.79	94.91
	Namibia	**500**	**7**	**0.4739785**	**3.59**	**95.79**
EA_2016_H	Khomas	1000	8	0.6040675	110.40	94.90
		500	6	0.6223167	112.06	94.75
	Namibia	**500**	**6**	**0.4705776**	**3.60**	**95.75**
EA_2011_true	Khomas	1000	7	0.6172502	87.84	95.11
		500	7	0.5979264	86.46	95.27
	Namibia	**500**	**6**	**0.4651136**	**3.54**	**95.92**
EA_2011_L	Khomas	1000	7	0.6020828	90.90	95.20
		500	6	0.6107100	91.55	95.13
	Namibia	**500**	**7**	**0.4669534**	**3.56**	**95.89**
EA_2011_M	Khomas	1000	8	0.6021614	99.72	95.14
		500	6	0.6264232	101.71	94.94
	Namibia	**500**	**7**	**0.4688343**	**3.58**	**95.83**
EA_2011_H	Khomas	1000	6	0.6018100	117.18	94.96
		500	7	0.5960483	116.62	95.01
	Namibia	**500**	**8**	**0.4712438**	**3.61**	**95.73**
EA_2006_true	Khomas	1000	7	0.7476114	120.25	94.39
		500	7	0.7438385	119.94	94.42
	Namibia	**500**	**7**	**0.4919520**	**3.68**	**95.61**
EA_2006_L	Khomas	1000	7	0.7370870	124.38	94.43
		500	8	0.7372104	124.39	94.43
	Namibia	**500**	**7**	**0.4938965**	**3.69**	**95.57**
EA_2006_M	Khomas	1000	6	0.6757099	128.83	94.80
		500	7	0.6671515	128.01	94.87
	Namibia	**500**	**7**	**0.4841646**	**3.66**	**95.63**
EA_2006_H	Khomas	1000	7	0.6889540	149.00	94.55
		500	8	0.6949817	149.65	94.51
	Namibia	**500**	**8**	**0.4902797**	**3.70**	**95.50**
EA_2001_true	Khomas	1000	9	0.7004286	142.87	94.87
		500	8	0.6985975	142.68	94.89
	Namibia	**500**	**8**	**0.4948715**	**3.71**	**95.51**
EA_2001_L	Khomas	1000	10	0.6891912	147.31	94.93
		500	8	0.6938707	147.81	94.90
	Namibia	**500**	**8**	**0.4976646**	**3.73**	**95.47**
EA_2001_M	Khomas	1000	6	0.7014662	159.95	94.80
		500	6	0.7094756	160.86	94.74
	Namibia	**500**	**7**	**0.4948384**	**3.73**	**95.47**
EA_2001_H	Khomas	1000	6	0.6953790	180.71	94.69
		500	6	0.6913627	180.19	94.72
	Namibia	**500**	**7**	**0.4915084**	**3.73**	**95.44**

Table 20. Random Forest model fit statistics for constituency-level input census scenarios

Scenario model	Coverage	Number of trees	Number of splits	Mean Square Error (MSE)	Coefficient of variation (CV)*	Percent variance explained
Const_2016_true	Namibia	500	5	0.3647471	3.12	94.76
Const_2016_L	Namibia	500	6	0.3527696	3.08	94.91
Const_2016_M	Namibia	500	5	0.3550769	3.10	94.83
Const_2016_H	Namibia	500	6	0.3713581	3.20	94.49
Const_2011_true	Namibia	500	5	0.3647895	3.14	94.73
Const_2011_L	Namibia	500	6	0.3545479	3.10	94.85
Const_2011_M	Namibia	500	5	0.3408526	3.05	95.01
Const_2011_H	Namibia	500	5	0.3573817	3.14	94.67
Const_2006_true	Namibia	500	6	0.3594167	3.14	94.70
Const_2006_L	Namibia	500	6	0.3650124	3.17	94.59
Const_2006_M	Namibia	500	5	0.3667935	3.19	94.52
Const_2006_H	Namibia	500	6	0.3500677	3.13	94.69
Const_2001_true	Namibia	500	6	0.3750283	3.23	94.33
Const_2001_L	Namibia	500	5	0.3532104	3.14	94.64
Const_2001_M	Namibia	500	5	0.3616736	3.19	94.47
Const_2001_H	Namibia	500	7	0.3623336	3.20	94.39

** multiplied by 10,000,000*

Table 21. Covariate importance scores for select final gridded population models

Covariate	EA 2016 True	EA 2011 Low	EA 2006 Med	EA 2001 High	Const. 2016 True	Const. 2011 Low	Const. 2006 Med	Const. 2001 High
cov_011	1,565.7	1,466.3	1,468.4	1,487.0	12.0	13.8	15.2	13.3
cov_040	399.7	321.0	304.3	344.0	14.7	12.2	14.5	14.4
cov_130	208.2	189.6	178.4	201.3	2.8	1.8	2.4	2.2
cov_140	493.1	311.2	290.7	449.4	2.4	1.7	2.5	1.8
cov_150	457.0	422.9	370.7	356.6	4.0	2.8	3.7	2.6
cov_160	446.2	414.1	401.4	370.7	3.8	2.9	3.2	2.8
cov_190	5,710.1	5,200.9	5,550.0	5,258.2	98.0	102.0	110.9	95.2
cov_200	466.6	338.9	362.0	365.0	3.5	2.6	5.0	2.2
cov_bsgme	11,929.0	13,650.1	13,433.1	13,855.7	148.4	167.9	136.6	185.6
cov_cciwat	1,152.3	688.8	799.8	928.1	47.2	45.0	49.8	42.6
cov_coast	685.9	532.5	500.8	571.8	6.3	4.4	5.6	3.1
cov_guf	14,885.8	18,960.6	16,996.1	14,547.7	77.2	74.3	90.2	68.7
cov_intsec	2,928.9	1,802.7	1,674.7	1,939.7	144.9	161.6	128.9	148.8
cov_prec	258.2	222.7	245.1	251.9	55.7	45.6	44.6	30.0
cov_road	5,518.9	3,094.2	4,092.4	4,484.8	1.6	1.6	1.8	1.2
cov_slope	557.3	516.3	533.2	624.6	58.2	45.1	44.7	39.8
cov_topo	1,535.2	1,419.0	1,437.7	1,519.2	7.3	7.3	7.7	6.2
cov_tt50k	1,127.7	660.2	619.5	692.8	12.0	8.5	8.5	2.8
cov_viirs	10,469.3	10,369.6	8,985.0	7,778.2	8.8	9.6	11.4	4.7
cov_waterw	814.6	664.9	596.6	623.2	24.7	13.7	22.6	8.0
cov_wdpa	935.1	717.4	739.7	835.5	3.5	2.0	2.0	2.2
cov_011	1,565.7	1,466.3	1,468.4	1,487.0	12.0	13.8	15.2	13.3
cov_040	399.7	321.0	304.3	344.0	14.7	12.2	14.5	14.4
cov_130	208.2	189.6	178.4	201.3	2.8	1.8	2.4	2.2

4.4 Cell-level accuracy of simulated gridded population sample frames versus simulated "real" populations

4.4.1 Methods

No accuracy assessments have been performed on gridded population data at the cell-level; accuracy has always been assessed at the scale of the input population data (see section 4.2.1.5). Using the simulated "real" population household locations, this analysis provides a rare opportunity to assess Random Forest model accuracy at the scale of the output grid cell. This section summarizes the statistics used to make comparisons in grid cells that ranged in size from 100 metres by 100 metres to 1 kilometre by 1 kilometre. All analyses were performed in R 3.5.2 on the Iridis 5 high performance computing cluster.

To evaluate cell-level accuracy, I used two error metrics, RMSE and MAE (both detailed in 4.2.1.5), to understand the degree of inaccuracy at the cell-level. I additionally created two usability measures with respect to maximum error in both the population estimate and the estimated population distribution. Accuracy of the estimated population distribution (percent of overall estimated population in each cell) is important for drawing an accurate sample and calculating accurate sampling probabilities. However, accuracy in the total population estimate is also important for field operations. A sampling unit with too few people wastes resources and potentially decreases sample size, while a sampling unit with too many people requires additional resources and time to perform segmentation (i.e., manually splitting the sampling units during fieldwork).

To measure accuracy of the population distribution, I calculated the percent of cells whose estimated population was +/- 20% the "true" population (PerDiff20). Given that a typical EA has 200 households and ranges from 100 to 300 households, 20% larger than 300 households would be 360 households, roughly the point at which an EA would need to be segmented in survey practice. To generate this statistic, I first calculated percent difference between the estimated and "true" populations in each cell i.

$$PerDiff_i = \frac{|\hat{y}_i - y_i|}{y_i} \times 100$$

To be able to measure percent difference, I added 0.000001 to every cell in the estimated population and the "true" population to have non zero values in the denominator; this amounted to 4 people being added to each dataset for all of Khomas. Then I classified each cell as 0 if it was greater than 20%, or 1 if it was less than or equal to 20%. Finally, I calculated the percentage of n cells less than or equal to 20% of the "true" population distribution.

$$PerDiff20 = \frac{n(\leq 20)}{n} \times 100$$

I additionally calculated the percent of cells that had an absolute difference of 10 people (AbsDiff10) or 25 people (AbsDiff25) per hectare from the "true" population count using a similar approach. In Khomas, where the average household size is 3.7 people (NSA, 2011b), this would amount to +/- 3 or 7 households per hectare. First I calculated the absolute difference between the estimated and true populations in cell i.

$$AbsDiff_i = |\hat{y}_i - y_i|$$

Then I classified cells as 0 if the difference was greater than 10 (or 25) people per hectare, and 1 if the difference was 10 (or 25) people and fewer. To calculate people per hectare, this number was increased by the same factor f^2 as cell size (e.g., factor = 2 for 200 metre by 200 metre cell). Then I calculated the percent of n cells with an estimated population less than or equal to that value.

$$AbsDiff10 = \frac{n(\leq 10 \times f^2)}{n}$$

$$AbsDiff25 = \frac{n(\leq 25 \times f^2)}{n}$$

Given that the WorldPop-RF / WorldPop-Global model attribute a fraction of a person to each cell in unsettled areas, I ran these statistics on the full estimated population for each scenario, as well as a version of the dataset in which 100 metre by 100 metre cells with a population estimate less than 1 were excluded, and another version in which cells with an estimated population less than 4 (approximately 1 household) were excluded. I also stratified the analysis by urban and rural cells using my earlier classification of EAs in 2016 (see section 3.5.3) to evaluate whether the model performed differently by urbanicity.

4.4.2 Results

I first reviewed a summary of all statistics across all 16 scenarios and 10 cell sizes to evaluate whether stratification by urbanicity was necessary, and whether I would present results with any population exclusion. The statistics were highly influenced by near zero population estimates in cells located in Khomas's vast unsettled areas. Whereas the "real" population covered a total of 11,206 100 metre by 100 metre cells (not reported), the estimated population covered 4,672,293 rural cells, and 14,403 urban cells (Table 22). Millions of near zero cell-level estimates in the gridded population severely skewed MAE (3), AbsDiff10 (99.8%), and AbsDiff25 (99.9%) to appear highly accurate because so many cells with near zero population estimates were compared to a "true" population of zero (Table 22).

After excluding cell estimates less than 1 or 4, the same statistics were far more reasonable (MAE: 483 and 656; AbsDiff10: 69.8% and 60.0%; AbsDiff25: 84.3% and 78.2%) (Table 22). Visual inspection of the 100 metre by 100 metre gridded population dataset derived from 2016 "true" population counts aggregated to EA showed that a substantial number of "true" population would be omitted from the analysis if I excluded cells with a population estimate less than 4 (Figure 25), but only a few people in the "true" population would be omitted if I excluded cell population estimates less than 1 (Figure 26), thus all results are presented for areas with 1 or more estimated people in the original 100 metre by 100 metre model.

The results revealed marked urban-rural differences in cell-level accuracy. For example, in cells with 1+ estimated persons, the average accuracy statistics across datasets and scales were: RMSE (U:1,506 vs. R:185), MAE (U:987 vs. R:64), AbsDiff10 (U:44.0% vs R:92.1%), AbsDiff25 (U:68.6% vs R:97.9%), and PerDiff20 (U:77.9% vs R:18.4%) (Table 22). Thus, all further results are stratified by urbanicity.

Table 22. Summary of accuracy statistics across all simulated gridded population datasets and grid cell sizes, by urban/rural and exclusion of cells with small estimated population

Red indicates the results presented in detail in this chapter

	100m	RMSE	MAE	AbsDiff10		AbsDiff25		PerDiff20	
	N cells	Mean	Mean	Mean	Range	Mean	Range	Mean	Range
No exclusion									
All	4,686,696	78	3	99.8%	(99.7, 99.9)	99.9%	(99.8, 99.9)	2.1%	(0.0, 6.6)
Rural	4,672,293	11	0	100.0%	(99.8, 100.0)	100.0%	(99.9, 100.0)	1.8%	(0.0, 6.3)
Urban	14,403	1,405	874	50.8%	(45.1, 66.3)	72.3%	(68.8, 85.8)	74.9%	(3.1, 97.9)
Exclude cells with estimated population < 1									
All	25,746	1,024	483	69.8%	(39.8, 76.3)	84.3%	(71.6, 87.8)	45.3%	(2.0, 78.6)
Rural	13,861	185	64	92.1%	(24.3, 99.3)	97.9%	(64.4, 100.0)	18.4%	(0.1, 54.0)
Urban	11,885	1,506	987	44.0%	(35.8, 60.0)	68.6%	(63.2, 83.1)	77.9%	(3.5, 100.0)
Exclude cells with estimated population < 4									
All	21,824	1,209	656	60.0%	(31.1, 72.5)	78.2%	(63.0, 86.2)	50.3%	(2.9, 93.4)
Rural	12,333	238	98	88.5%	(17.4, 99.4)	97.2%	(61.1, 100.0)	20.8%	(0.0, 78.6)
Urban	9,491	1,665	1,176	34.1%	(21.7, 51.4)	61.9%	(50.0, 79.2)	80.9%	(4.0, 100.0)

Below I visualise each statistic for EA and constituency on the same scale so that relative differences are easier to spot across the different simulated census scenarios and aggregated grid cell sizes. I present results for absolute difference of +/- 10 people per hectare (AbsDiff10), rather than +/- 25 people (AbsDiff25), because the larger value would represent huge errors in the estimated population in large grid cells. In 1 kilometre by 1 kilometre cells, for example, +/- 10 people per hectare would result in +/- 1000 people overall, while +/- 25 people per hectare would result in +/- 2500 people.

A summary of RMSE is presented in Figure 27, a summary of MAE is presented in Figure 28, and AbsDiff10 and PerDiff20 are visualized together in Figure 29. I visualized absolute and percent population differences together because they are both important to survey fieldwork, and I wanted to make recommendations of for a minimum grid cell size to use in the next analysis of gridded population sampling with both in mind. All statistics are presented in full in Appendix E.

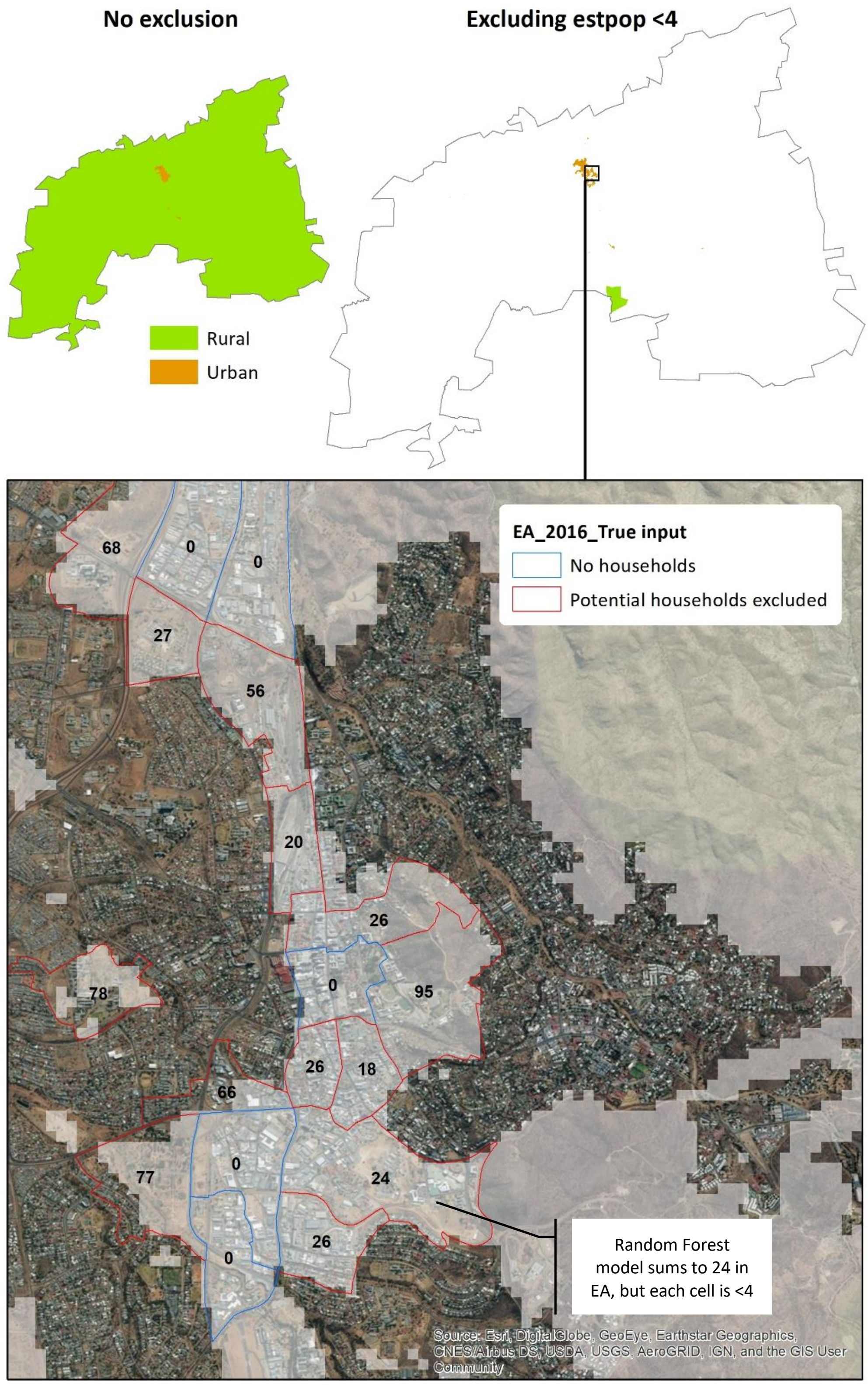


Figure 25. Visual of areas included in the analysis, and select EAs with "true" population, when cells with an estimated population <4 are excluded (EA_2016_true 100m dataset)

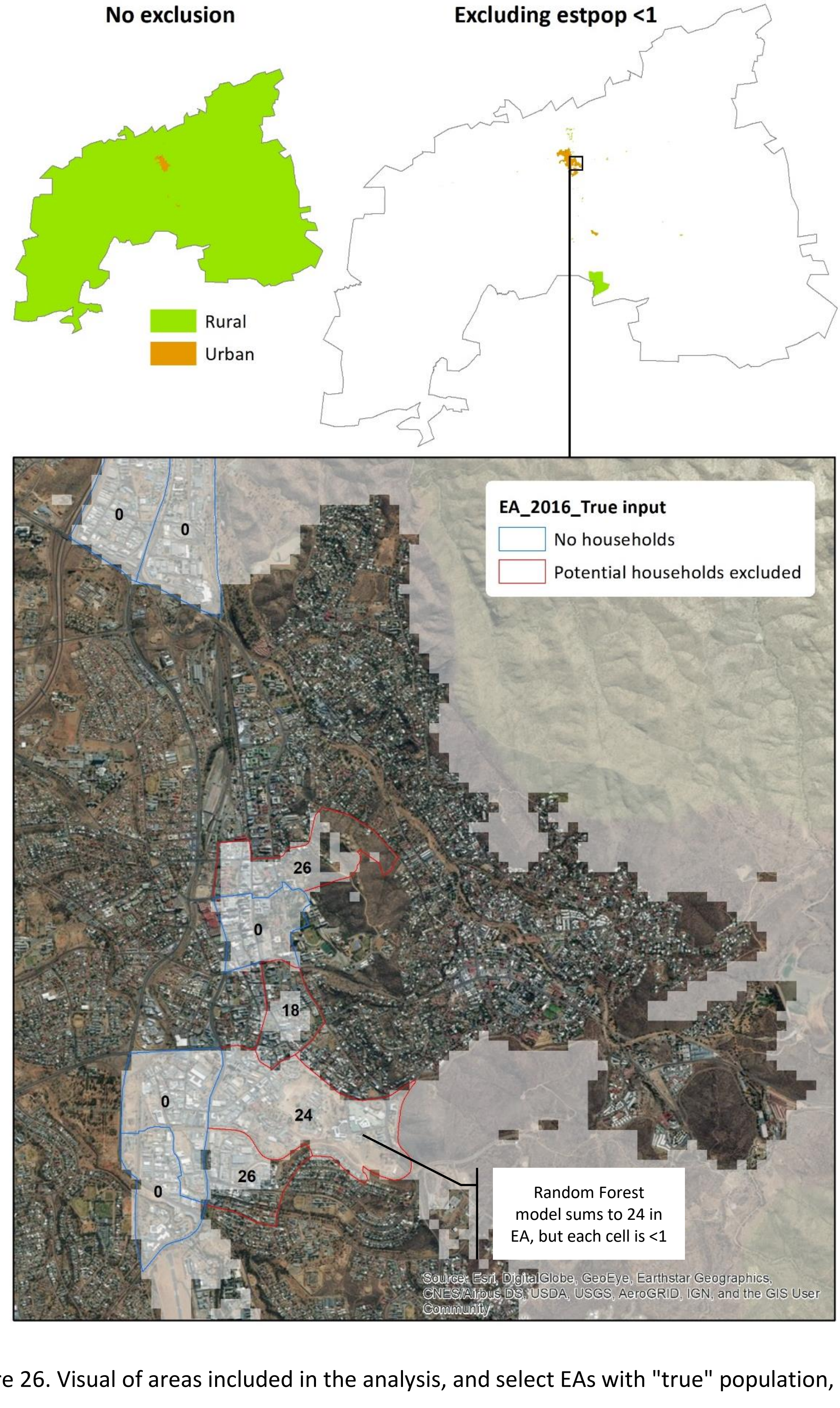


Figure 26. Visual of areas included in the analysis, and select EAs with "true" population, when cells with an estimated population <1 are excluded (EA_2016_true 100m dataset)

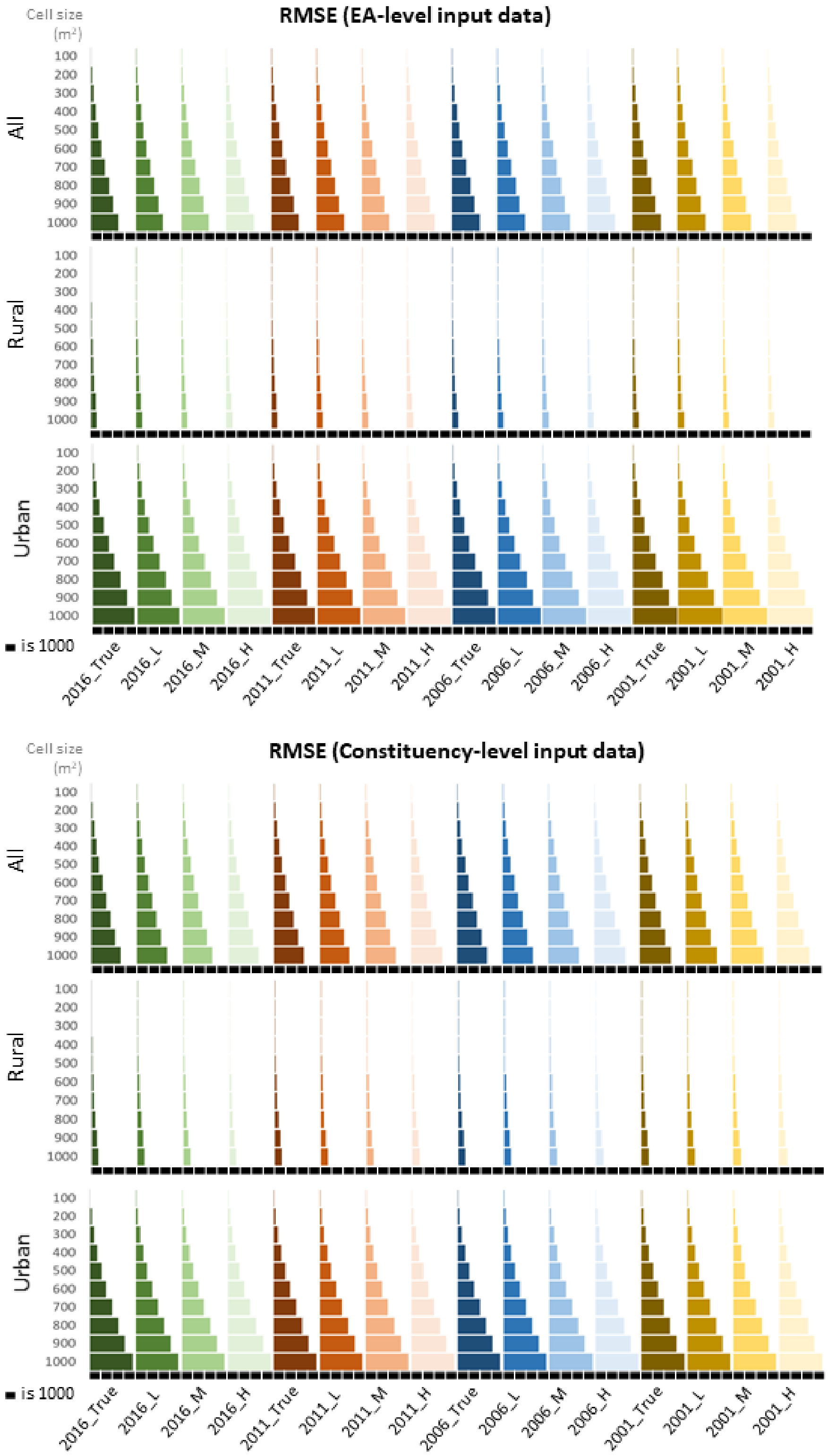


Figure 27. RMSE in gridded population cells, by grid cell size, simulated census scenario, and aggregation of input population

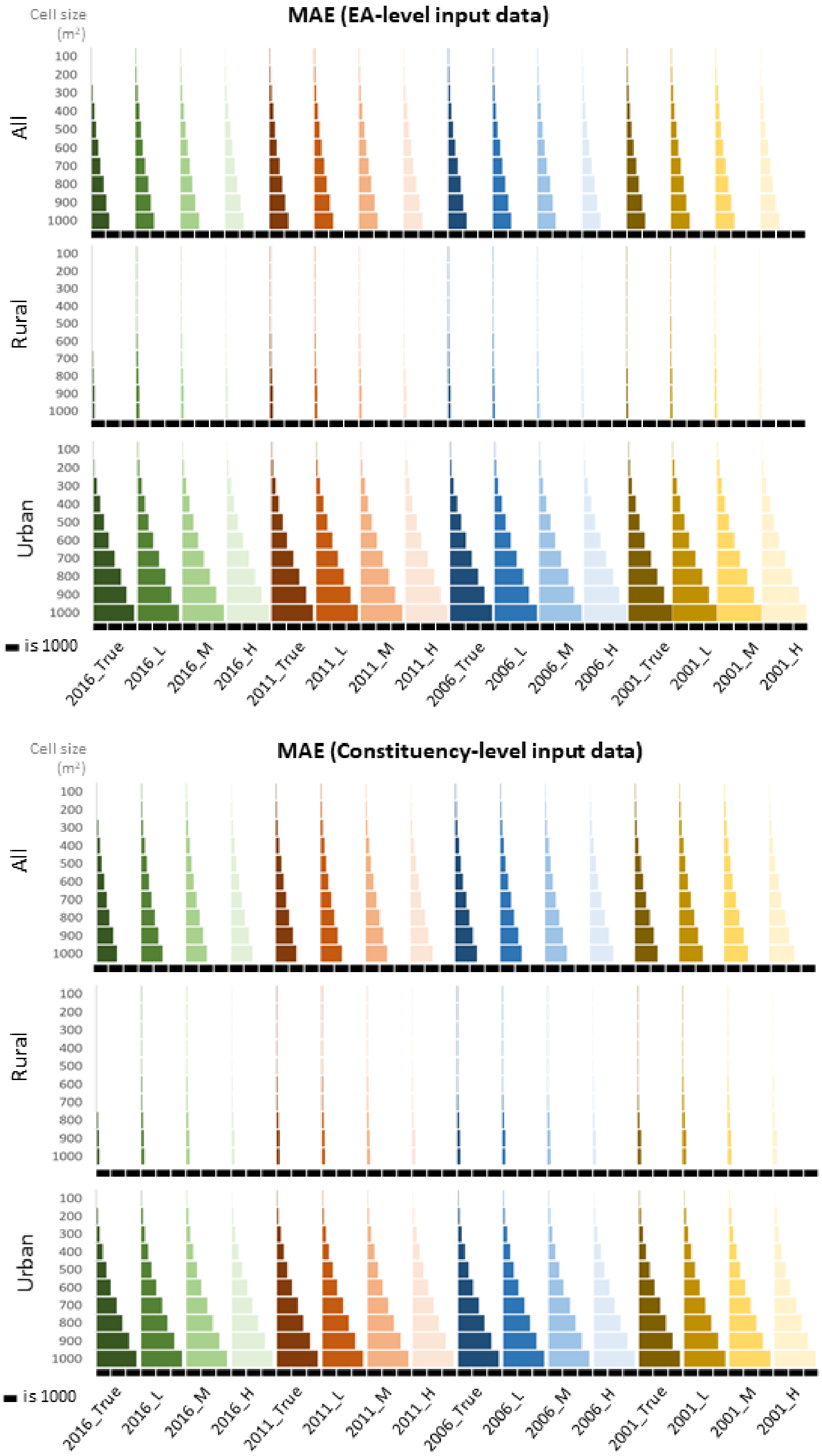
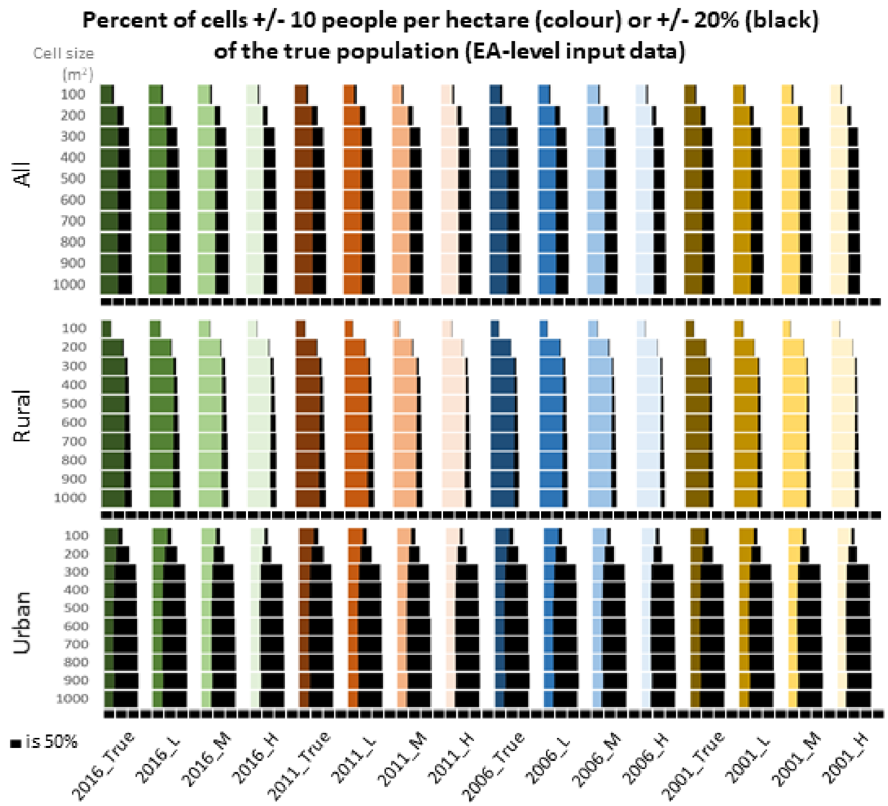


Figure 28. MAE in gridded population cells, by grid cell size, simulated census scenario, and aggregation of input population



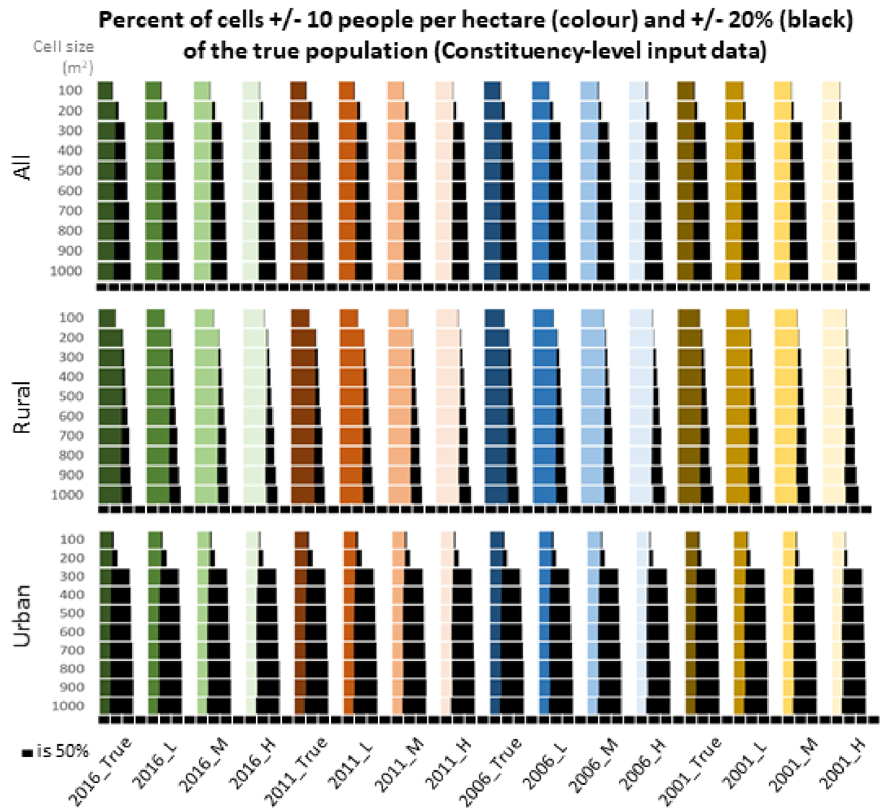


Figure 29. Percent of cells +/- 10 people per hectare (AbsDiff10) or +/- 20% of the true population distribution (PerDiff20), by grid cell size, simulated census scenario, and aggregation of input population

RMSE and MAE in grid cells did not differ substantially across the simulated census scenarios which included varying degrees of population missingness and outdatedness (Figure 27, Figure 28). Furthermore, errors only slightly decreased when the input data were aggregated to EA rather than constituency. The major driver of RMSE and MAE in cells was urban location; error in urban cell-level estimates was substantially higher than rural cells according to both metrics (Figure 27, Figure 28). The errors increased linearly with cell size, likely because large errors in urban areas were compounded as cell size increased; RMSE and MAE only increased slightly in rural areas with cell size.

At the 100 metre by 100 metre scale, few grid cells in urban or rural areas were accurately estimated to have +/- 10 people when compared to the "true" population (Figure 29). However, this quickly increased in rural grid cells but decreased in urban grid cells when grid cells were aggregated. At 300 metres by 300 metres, there was a sharp increase in cells that had +/- 20% the true population in urban areas (PerDiff20), with smaller increases in rural areas (Figure 29). In both urban and rural areas, the number of cells with +/- 20% the true population (PerDiff20) remained fairly constant when cells were aggregated to cells larger than 300 metres by 300 metres. As with RMSE and MAE, the cell-level statistics for AbsDiff10 and PerDiff20 were dominated by urban/rural differences, rather than outdatedness or inaccuracies in the input population, or level of input population aggregation.

At 300 metres by 300 metres, most estimates in rural cells were +/- 10 people per hectare and most estimates in urban cells were +/- 20% the true population. To demonstrate that differences due to input population outdated and inaccuracy were subtle or negligible, Table 23 presents statistics for select census scenarios in 300 metre by 300 metre cells. There were only slight increases in RMSE and MAE as input datasets were increasingly outdated and inaccurate, and when the input data were aggregated. For example, RMSE was 238 for EA_2016_True (most accurate, fine scale input) and 286 for Const_2001_H (least accurate, coarse input). AbsDiff10 and AbsPer20 were negligible or decreased very slightly in scenarios of greater outdatedness or inaccuracy, but only when the input data were at the EA scale (Table 23). There was an unexpected improvement in PerDiff20 in scenarios of greater outdatedness and inaccuracy among input datasets aggregated to constituency (from 41.4% to 49.2%), likely because the incorrect low constituency population counts approached the underestimated population produced by the Random Forest model (Table 23).

Table 23. Summary of accuracy statistics in 300 metre by 300 metre grid cells for select simulated census scenarios, excluding areas with an original estimated population less than 1

	RMSE			MAE			AbsDiff10			PerDiff20		
	All	R	U	All	R	U	All	R	U	All	R	U
EA												
2016_True	238	36	353	112.9	22.1	223.6	73.5	97.3	44.7	44.6	10.8	86.6
2011_L	245	40	365	118.1	25.5	233.0	72.6	96.4	43.4	43.6	9.1	87.1
2006_M	245	35	374	114.2	22.7	237.9	73.7	97.7	41.6	41.5	7.2	88.6
2001_H	254	36	394	117.6	23.0	252.5	73.4	97.8	39.0	39.7	4.6	90.3
Constituency												
2016_True	253	25	361	111.6	10.0	218.0	74.3	98.7	48.7	41.4	8.8	75.6
2011_L	258	25	365	114.2	9.4	221.3	73.6	98.5	48.2	42.2	9.1	76.0
2006_M	266	26	372	119.0	8.1	226.5	72.4	98.3	47.3	43.6	9.5	76.7
2001_H	286	29	379	135.8	8.5	232.1	68.4	97.7	46.3	49.2	11.8	77.6

To determine the magnitude and potential sources of the massive cell-level errors, particularly in urban areas, I performed two sub-analyses. I calculated cell-level bias, the average difference between the “true” population and the estimated population, and percent of the total population in Khomas that was misallocated to cells which were unsettled according to the “true” population.

In the 100 metre by 100 metre estimates derived from “true” 2016 census data, cell-level bias ranged from 245 population underestimate to 525 population overestimate with EA-level input, and from 333 population underestimate to 96 population overestimate with constituency-level input (unreported). A visual comparison of bias is presented in Figure 30 for a select EA and constituency, demonstrating that within-urban cell-level estimates based on the most accurate data available differed substantially from reality.

The average 300 metre by 300 metre urban cell underestimated the population by more than 200 people, while the average rural cell was underestimated by 3 (constituency-level input) to 14 (EA-level input) people (Table 24). In this analysis, where I excluded millions of cells in unsettled areas with a fraction of an estimated person, both urban and rural cell-level estimates suffered greater bias when grid cells were aggregated. The average 1 kilometre by 1 kilometre urban cell was missing approximately 2,400 people, while the average rural cell of the same size was missing approximately 140 people (Table 24). When these same statistics were performed on datasets without exclusions, the bias in rural areas was close to zero due to the enormous number of near zero estimates compared to zero “true” population, while the bias in urban areas was nearly identical (unreported).

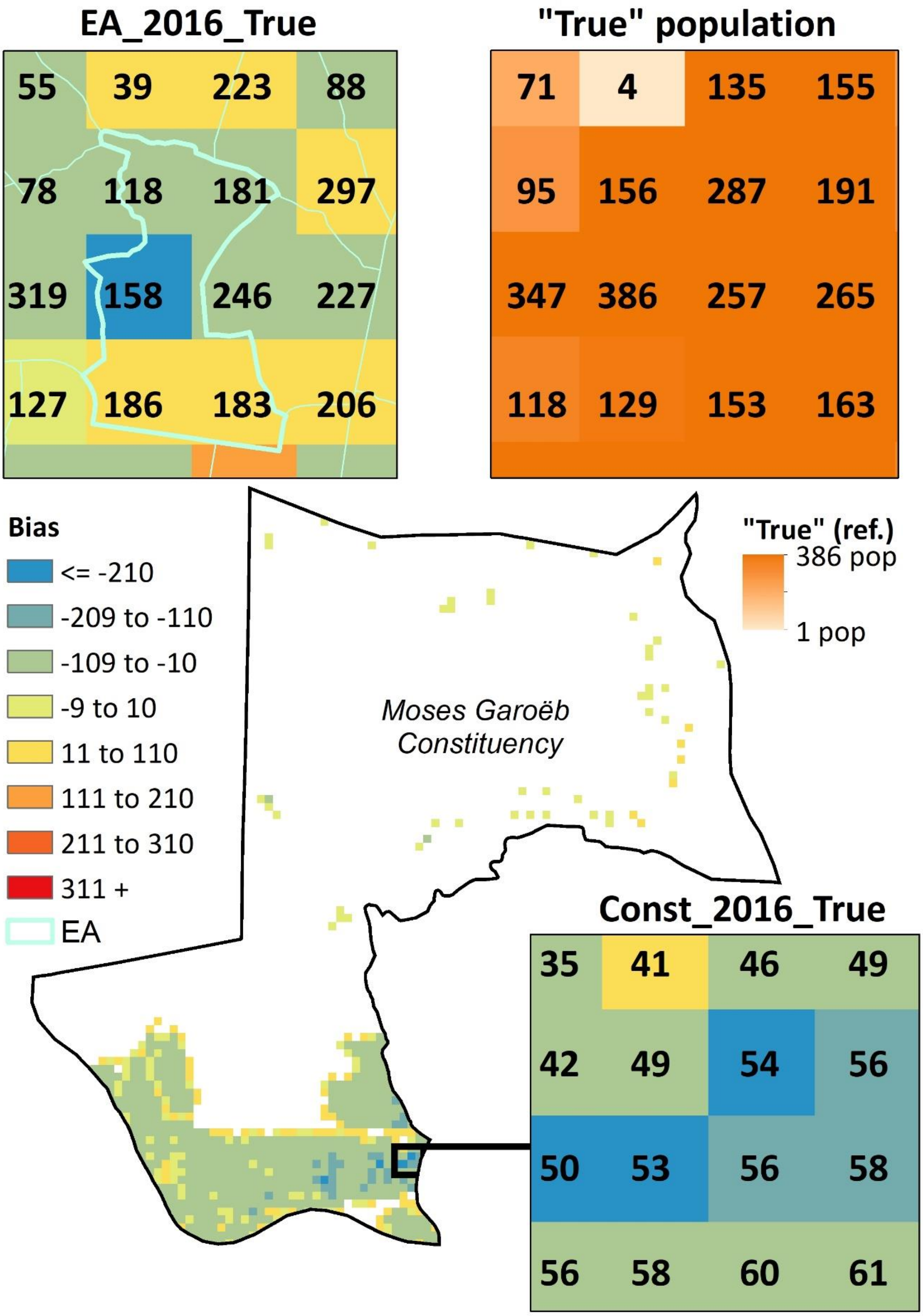


Figure 30. Example 100 metre by 100 metre gridded population estimate bias versus "true" population in a select EA and constituency

Table 24. Bias in the 2016 gridded population estimates derived from “true” population counts, by grid cell size and aggregation level of the input data, excluding areas with an original estimated population less than 1

Grid cell size (m^2)	EA_2016_true			Const_2016_true		
	All	Rural	Urban	All	Rural	Urban
100	10	20	0	0	7	-8
200	-29	18	-85	-39	6	-86
300	-92	14	-223	-103	3	-214
400	-177	8	-416	-192	-1	-394
500	-283	3	-650	-306	-8	-616
600	-407	-22	-933	-445	-34	-891
700	-551	-33	-1,293	-611	-51	-1,229
800	-718	-72	-1,664	-787	-90	-1,556
900	-891	-126	-2,026	-994	-152	-1,974
1000	-1,073	-126	-2,476	-1,233	-167	-2,421

Table 25 summarises the percent of the estimated population misallocated to “true” unsettled cells. The exclusion of cells with fewer than one estimated person was lifted for this analysis to understand error in the underlying estimates. Roughly 20% (EA-level input) or 10% (constituency-level input) of the population was misallocated to unsettled 100 metre by 100 metre cells (Table 25). However, as cells were aggregated, the percent of misallocated population dropped precipitously. For example, at 300 metres by 300 metres, less than 3% (EA-level input) or 1% (constituency-level input) of Khomas’s population was misallocated to unsettled cells. This indicated that most of the population was disaggregated within, or near to, settlements. The rates of misallocation were similar across grid cell sizes when cells with less than one person were excluded (not reported).

Table 25. Percent of the overall population (no exclusion) that is misallocated to unsettled cells, by simulated census scenario, aggregation level of the input data, and grid cell size

	Grid cell size (m²)									
	100	200	300	400	500	600	700	800	900	1000
Scenario	%	%	%	%	%	%	%	%	%	%
EA_2001_H	20.05	4.95	2.22	1.28	0.81	0.59	0.41	0.32	0.26	0.21
EA_2001_L	21.15	5.21	2.33	1.35	0.86	0.62	0.43	0.34	0.27	0.22
EA_2001_M	20.63	5.08	2.27	1.31	0.83	0.60	0.42	0.33	0.26	0.21
EA_2001_true	21.90	5.40	2.42	1.40	0.89	0.64	0.45	0.35	0.28	0.23
EA_2006_H	20.98	5.15	2.30	1.32	0.84	0.61	0.43	0.33	0.26	0.22
EA_2006_L	22.68	5.57	2.49	1.44	0.91	0.66	0.46	0.36	0.29	0.24
EA_2006_M	20.47	5.00	2.23	1.29	0.82	0.59	0.41	0.32	0.26	0.21
EA_2006_true	23.16	5.69	2.54	1.47	0.93	0.67	0.47	0.37	0.29	0.24
EA_2011_H	20.63	5.03	2.24	1.29	0.82	0.59	0.41	0.33	0.26	0.21
EA_2011_L	23.34	5.68	2.53	1.46	0.93	0.66	0.47	0.37	0.29	0.24
EA_2011_M	26.50	6.50	2.90	1.67	1.06	0.76	0.54	0.42	0.33	0.27
EA_2011_true	21.67	5.26	2.34	1.35	0.85	0.62	0.43	0.34	0.27	0.22
EA_2016_H	20.29	4.94	2.20	1.27	0.81	0.58	0.41	0.32	0.25	0.21
EA_2016_L	22.00	5.34	2.38	1.37	0.87	0.63	0.44	0.35	0.27	0.23
EA_2016_M	18.69	4.50	2.00	1.15	0.73	0.52	0.37	0.29	0.23	0.19
EA_2016_true	20.82	5.02	2.23	1.28	0.81	0.59	0.41	0.32	0.26	0.21
const_2001_H	5.80	1.19	0.46	0.23	0.13	0.08	0.06	0.04	0.03	0.02
const_2001_L	7.33	1.53	0.60	0.30	0.17	0.11	0.07	0.05	0.04	0.03
const_2001_M	6.77	1.40	0.54	0.28	0.16	0.10	0.07	0.05	0.04	0.02
const_2001_true	7.59	1.58	0.61	0.31	0.18	0.11	0.07	0.05	0.04	0.03
const_2006_H	7.10	1.46	0.56	0.29	0.16	0.10	0.07	0.05	0.04	0.03
const_2006_L	8.78	1.83	0.71	0.36	0.21	0.13	0.08	0.06	0.04	0.03
const_2006_M	8.13	1.69	0.66	0.34	0.19	0.12	0.08	0.06	0.04	0.03
const_2006_true	9.22	1.93	0.75	0.38	0.22	0.14	0.09	0.06	0.05	0.03
const_2011_H	8.35	1.73	0.67	0.34	0.19	0.12	0.08	0.06	0.04	0.03
const_2011_L	10.86	2.28	0.89	0.46	0.26	0.16	0.11	0.08	0.06	0.04
const_2011_M	9.79	2.04	0.79	0.41	0.23	0.15	0.09	0.07	0.05	0.03
const_2011_true	11.34	2.39	0.94	0.48	0.27	0.17	0.11	0.08	0.06	0.04
const_2016_H	8.95	1.87	0.73	0.37	0.21	0.13	0.09	0.07	0.05	0.03
const_2016_L	11.66	2.45	0.96	0.49	0.28	0.18	0.11	0.08	0.06	0.04
const_2016_M	10.69	2.25	0.88	0.45	0.26	0.16	0.11	0.08	0.06	0.04
const_2016_true	12.47	2.63	1.03	0.53	0.30	0.19	0.12	0.09	0.06	0.05

4.5 Discussion

4.5.1 Accuracy of simulated gridded population sample frames

Urban/rural population difference in Khomas dominated the results. While I found clear evidence of my hypothesis that the accuracy of gridded population estimates can be improved by aggregating grid cells, there was limited evidence in this context that outdated, inaccurate census data plays a major role in the cell-level accuracy of gridded population estimates using a WorldPop-RF / WorldPop-Global modelling approach. Three potential issues might explain cell-level inaccuracies:

(1) non-zero population estimates were attributed to millions of unsettled cells;
(2) presence of local spatial inaccuracies due to the coarse resolution of some covariates and/or a weak relationship between population density and the covariates; and
(3) the assumption that average EA- or constituency-level population densities per hectare provide a sufficient range of population values from which to model population density at the 100 metre by 100 metre (hectare) scale.

The first issue, non-zero population estimates in millions of unsettled cells, probably played only a minor role in cell-level inaccuracies. Table 25 demonstrates that even in this context of vast unsettled areas, only a small portion of Khomas's population was misallocated to cells far from actual settlements. Nearly all of the population was estimated to be in cells within 200 to 300 metres of the "true" population.

The second issue, related to covariate resolution and the relationship of covariates with population density, probably played a larger role. A number of the Random Forest model covariates, such a land cover type and night-time lights, had an original resolution substantially larger than 100 metre by 100 metres, which could have resulted in a halo effect around settlements, causing populations to be disaggregated to cells near a settlement, but not directly over it. Figure 29 and Table 25 provide evidence of this; the accuracy of the estimated population distribution, and correct allocation of population to settled cells, both performed well when the estimated population was aggregated to 300 metres by 300 metres or larger.

Other covariates, such as road and intersection locations, and urban or built-up settlement area, were available at very fine spatial resolution and thus were precise at the 100 metre by 100 metre scale. Although they are good indicators of a settlement, they are not necessarily good indicators of higher or lower population density within a settlement. The lack of fine-scale covariates associated with population density within cities and towns likely explains a portion of the cell-level error observed. Other issues that might further decrease local spatial accuracy are temporal miss-match of covariates (Leyk *et al.*, 2019) and covariate spatial autocorrelation (Sinha *et al.*, 2019).

The third potential source of cell-level inaccuracies, using average population densities from large administrative units to estimate population density in much smaller grid cells, is known as the ecological fallacy (Thorndike, 1939; Selvin, 1958), and probably played the largest role in cell-level inaccuracies. Population densities are used by the random forest model to establish relationships between covariates and population (total population divided by total area), not population totals. In the simulated "true" 2016 household point locations dataset, the maximum number of people in a 100 metre by 100 metre cell was 386 people. However, in the same 100 metre by 100 metre grid, the maximum population estimated by the Random Forest model was 469 people (EA-level input) and 112 (constituency-level input). Even with perfect covariates and exclusion of unsettled areas, this means that cells with high population density might be severely underestimated, and that the excess population will be disaggregated to other less dense cells in the same input administrative unit due to dasymetric reweighting.

The ecological fallacy was likely the main source of cell-level error in the WorldPop-RF / WorldPop-Global model results.

Population density per hectare in administrative areas did not reflect population density at the cell-level.

The likely underestimation of cell-level high population density in existing WorldPop-RF / WorldPop-Global datasets gives me pause about the value of gridded population sampling to ensure coverage of slum dwellers in household surveys. A defining characteristics of LMIC slum areas is high population density (Nuissl and Heinrichs, 2013; Ezeh *et al.*, 2017; Mahabir *et al.*, 2018; Lilford *et al.*, 2019), though high-rise apartment buildings are another form of high density urban living. The relationship between high population density and socioeconomic status likely varies by city, and is an area that needs further study to understand the accuracy of gridded population estimates. Given that many of the highest density cells are likely located in slum areas, gridded population sampling from a WorldPop-RF / WorldPop-Global dataset might result in an under sample of slum dwellers.

4.5.2 Recommendations

I have three recommendations to potentially address sources of cell-level error in WorldPop-RF/ WorldPop-Global gridded population estimates. The first recommendation is in light of a high quality building footprint dataset that has only recently become available for multiple LMICs from Maxar, Digital Global, and Ecopia (Maxar, 2019). This group produces a building footprint layer by country, derived by extracting building boundary features from very high resolution satellite imagery. Each country dataset is reported to include 95% or more of actual buildings, including structures made with traditional materials (Maxar, 2019). These building footprint datasets are initially a commercial product, but are released freely and publicly after three years. Accurate building footprints are likely associated with population density within settlements and have a finer spatial resolution than 100 metres by 100 metres, making it a potentially powerful covariate to include in Random Forest models, and likely to improve within urban cell-level estimates. Gridded population data producers are already evaluating how to incorporate this dataset into gridded population models (D. Leasure, personal communication, 9 Feb 2020).

Although there are good arguments against using a modelled settlement layer to mask gridded population estimates because this tends to omit small rural settlements and remote households (Stevens *et al.*, 2015), use of the new Maxar, Digital Globe, Ecopia building footprints layer might be sufficiently accurate to do just this. If buildings made from traditional materials are reflected in the dataset, then building footprints could be used to create a 100 metre by 100 metre mask to constrain the Random Forest model, or to set Random Forest outputs to zero before dasymetrically reweighting with the input population counts. This would mean that only cells containing buildings would be attributed population, and that cell-level population estimates would still sum to the total population of the input unit. However, the use of building footprints as a covariate might not, alone, address cell-level inaccuracies within urban areas.

A third recommendation is to create 100 metre by 100 metre population training datasets that could be incorporated with other training data during the Random Forest modelling phase of WorldPop-RF / WorldPop-Global workflow. This would help to address the third issue driving cell-level inaccuracies, which is that Random Forest models generally do not have sensible population-per-hectare densities to predict population densities accurately in 100 metre by 100 metre cells within cities. The training datasets might be simulated, like the one I created in Chapter 3, or they might come from existing household survey enumerations. For example, World Bank LSMS surveys record household size and household latitude-longitude coordinate for every household in every sampling unit (i.e., cluster/PSU) during the mapping-listing activity before drawing a final sample of households (M. Wild, personal communication, July 2019). While household geolocations are quite sensitive data and are never shared, aggregation of population counts to 100 metre by 100 metre grids cells by the data collection team, would produce an anonymised, non-sensitive dataset. Even if these densities are only available for a sample of locations, they would provide the Random Forest model with more accurate maximum population values in 100 metre by 100 metre cells, as compared to administrative-level averages. Recent communication with the World Bank indicates their openness to producing and sharing this type of information from LSMS surveys (M. Wild, personal communication, July 2019).

Addressing the ecological fallacy in the WorldPop-RF / WorldPop-Global workflow by incorporating training data from smaller areas, and adding fine scale spatial covariates to the Random Forest model that are correlated with variation in urban population density, would together likely improve the accuracy of gridded population estimates and subsequent gridded population surveys. Further evaluation of these recommendations ought to be a priority for teams using the WorldPop-RF / WorldPop-Global modelling workflow.

4.5.3 Decision: Gridded population sample frame scale in this thesis

The analyses in this chapter indicate that aggregating existing WorldPop-RF / WorldPop-Global gridded population estimates to 300 metre by 300 metre grid cells or larger in a setting similar to Khomas, Namibia is the best available choice for fieldwork. Results indicate that areas which are at least 300 metres by 300 metres contain +/- 20% the true population distribution in urban areas, and +/- 10 people per hectare in rural areas. However, the results also indicate large errors in urban cell-level estimates, which could be addressed by stratifying samples or analyses along urban/rural boundaries. In the next chapter, I limit the analysis to urban Khomas, and choose a method to aggregate gridded population estimates into units larger than 300 metres by 300 metres for household survey sampling.

Chapter 5: Sampling

5.1 Overview

The accuracy of household survey sample data is dependent on the accuracy of the sample frame from which it was drawn. Biased estimates of indicators can have major economic and political implications if survey data are used for decision-making and resource allocation. Thus, this chapter explores the effect of sample frame choice on accuracy of survey results. Specifically, this chapter explores whether sampling from an outdated or inaccurate census sample frame is better, the same, or worse than sampling from a WorldPop-RF / WorldPop-Global gridded population dataset derived from that underlying outdated or inaccurate census.

5.1.1 Research Questions

Research Question C: Can a gridded population sample frame be used to select a representative sample of the population under realistic scenarios of outdated, inaccurate census input data?

Research Question D: Do gridded population samples more accurately represent poor and vulnerable households compared to equivalent census-based samples, under realistic scenarios of outdated, inaccurate censuses?

5.2 Background

5.2.1 Typical sampling: High income countries

The methods considered typical in household surveys today were initially established in high income countries (HICs) and later adopted in low- and middle-income countries (LMICs). The first recorded systematic collection of household-level data began at the end of the 18th century in England, Saxony, Prussia, Belgium, and the United States, and the earliest generalizations about household characteristics from these types of data began in the middle of the 19th century (Grosh and Glewwe, 2000). Is was not until the 1920s, however, that probability-based statistical theories, at the heart of modern household survey statistics, were developed.

Groves (2011) describes “three eras of survey research” that ensued in HICs (Groves, 2011). The first of which was the *era of invention* from 1930 to 1960 during which all of the basic methods and tools still used in household survey sampling were established. At the start of this era, Neyman (1934) published a seminal article about the use of probability-based sampling to generate unbiased estimates and measurable sampling errors. During this era, area-based sample frames (e.g., counties, census geographic units), stratification, and multistage sampling methods were all developed (Groves, 2011) and applied by governments in the US and Europe to monitor household welfare and poverty throughout the great depression and after World War II (Grosh and Glewwe, 2000; Groves *et al.*, 2009). Until 1960, most household surveys were paper-based, with face-to-face interviews conducted mainly by women or mailed questionnaires, and response rates were generally over 90% (Groves, 2011).

Groves (2011) describes the 1960s through the 1980s as the *era of expansion.* This is when computerized data processing become available to governments, academia, and the private sector, and when most private homes in HICs acquired a landline telephone (Groves, 2011). During the 1960s, mechanical punch-card readers made it possible to process paper-based survey responses rapidly, and Computer Assisted Telephone Interviewing (CATI) technologies along with random digit dialling made household surveys possible to implement at much reduced costs (Groves, 2011). As survey research evolved, cluster-based sampling rather than strata-based sampling became common place, and an improved understanding of response bias due to question wording and question order improved the quality of questionnaires (Groves, 2011).

During this era, governments across HICs invested in large-scale health, economic, and opinion surveys at home and abroad. Eurobarometer, a public opinion survey conducted across Europe, was established in 1973 consisting of approximately 1000 face-to-face interviews per country, supplemented with telephone surveys; nine Western European countries were initially covered and today the survey covers 27 European Union member countries (European Commission, 2008). In 1982, countries from across Europe began the process of harmonizing their national Labour Force Surveys, many of which were established in the 1960s (Eurostat, 2018). In 1984, the International Social Survey Programme was founded when four existing survey programmes from the US, Great Britain, Germany, and Australia merged; the ISSP survey is now conducted in some 57 countries across Europe, Asia, South America, Oceania plus South Africa using a range of survey modes (Skaarhoj, 2018). Also in 1984, the US Centers for Disease Control and Prevention launched the Behavioral Risk Factor Surveillance System survey in 15 US states to collect health information, and this annual phone survey continues today with coverage of all 50 US states and the District of Columbia (CDC, 2014). In addition to the explosion of household survey programmes in HICs, governments in North America and Europe began to sponsor large-scale household surveys in LMICs, which I will discuss next.

However, by the 1980s, typical survey methods in HICs began to pose new challenges. Foremost, phone-based surveys were becoming problematic as respondents grew less tolerant of unplanned calls and long questionnaires. Non-response rates and incomplete questionnaires became a new concern for survey researchers. The use of post-stratification weighting became common place to adjust for incomplete responses, and formal statistical models were proposed to adjust for incomplete coverage of sample frames and for large non-response rates (Rubin, 1987). Until this point, most surveys were cross-sectional, meaning that respondents were enrolled at one point in time. However, longitudinal surveys, meaning surveys that enrol respondents over a period of time and collect multiple survey measurements, were introduced during this period.

According to Groves (2011), the 1990s through the present represents the third era in survey research described as *"designed data" supplemented by "organic data."* In this era, face-to-face household surveys declined substantially in HICs, and are now limited mainly to the first wave of data collection in longitudinal surveys. Cross-sectional surveys now tend to rely on a mix of mail, phone, and internet-based samples, with internet surveys drawing from volunteer panels. None of these sample modes are particularly representative of the population. As mobile phones have become ubiquitous, landline phones represent an increasingly middle- and upper-class section of older residents, and surveys from mobile phone numbers make households difficult to track, because mobile phones are usually associated with an individual rather than a household (Groves, 2011). For these reasons, survey researchers have, in recent years, turned to the use of organic data, generated passively by the general population and collected via internet scraping, social media aggregation, and other methods, to supplement imperfectly designed survey data (Groves, 2011).

It was during the second era, the *era of expansion*, that HICs began to invest in large-scale survey programmes in LMICs, and the trajectory of survey research methods in HIC and LMIC contexts diverged. Throughout the 1970s, countries around the world prioritized investments in primary healthcare, specifically investing in basic maternal, reproductive, and child health at local health centres and via community health worker programmes (Rohde *et al.*, 2008). These efforts culminated in the signing of the Alma Ata declaration by 134 countries, 67 international organizations, and numerous non-governmental organizations in 1978. The declaration prioritized primary healthcare "based on practical, scientifically sound and socially acceptable methods and technology made universally accessible through people's full participation and at a cost that the community and country can afford" so that everyone globally would have access to basic healthcare by the year 2000 (Rohde *et al.*, 2008). This shaped the focus of LMIC surveys.

5.2.2 Typical sampling: Low- and middle-income countries

As HICs scaled up funding for development in LMICs, they also scaled up funding of nationally-representative household survey programmes to fill a gap in data to monitor health needs and progress. A timeline of all major demographic, health, and economic survey programmes in LMICs sponsored by HICs is presented in Figure 31. In the 1970s, few LMICs had fully functioning vital registration systems or data from health systems; furthermore, few LMICs had existing household surveys to fill this information gap. India was among the first and only LMICs to establish a nationally-representative household survey programme, which launched in 1950 and continues today (Katyal *et al.*, 2013). In 1972, the United Nations (UN), International Statistical Institute, and International Union for the Scientific Study of Population joined forces to launch the five-year World Fertility Surveys (WFS) in several LMICs (International Statistical Institute, 1973). The WFS used an area-based sample frame, two stages of sampling (small areas, then households), and interviewed women of reproductive age about fertility, child mortality, and family planning. In 1975, the US Centers for Disease Control and Prevention (CDC) and the US Agency for International Development (USAID) collaborated to launch the Contraceptive Prevalence Survey (CPS) programme using a similar sample design, which evolved into the Family Planning and Maternal Child Health Survey (FP-MCH) when additional questions about breastfeeding, immunizations, health, and mortality where included (Lewis, 1983).

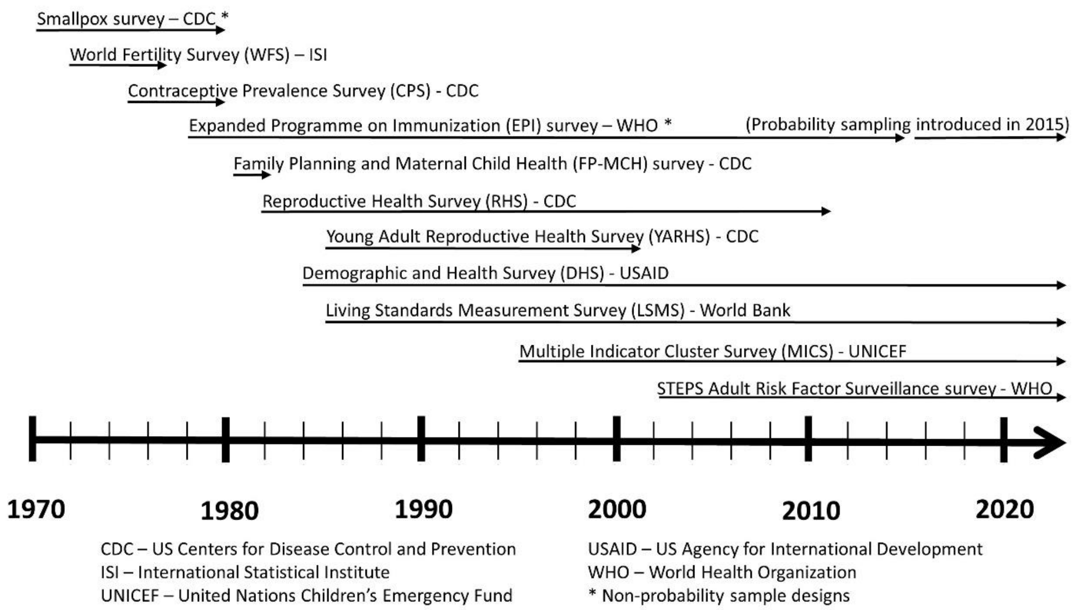


Figure 31. Timeline of household survey programmes in low- and middle-income countries

Meanwhile, the CDC and World Health Organization (WHO) each conducted household surveys focused on immunizations during smallpox and other vaccination campaigns starting in the 1970s (Morris, 2000). The vaccination coverage surveys, however, used non-probability sample designs, including the “30x7” quota design in which 30 small areas were sampled, and then seven households with young children were identified and interviewed using a random-walk or spin-the-pen method (Henderson and Sundaresan, 1982). In these quota designs, non-responding and absent households were not tracked, and then they were replaced with a neighbouring household preventing statistical adjustments for bias using sample weights. The further use of the random-walk or spin-the-pen methods, in which fieldworkers decided which households to sample, introduced potential bias from interviewers avoiding undesirable households (Grais, Rose and Guthmann, 2007; Cutts *et al.*, 2016). It would take years before strong critiques of these household survey methods took hold. Initially, household surveys in LMICs provided unprecedented information to international donors and national governments alike. Though, despite efforts, by 1985, less than 25% of all LMICs, and only about 6% of Sub-Saharan Africa’s population, had reliable data about household health or wellbeing (Grosh and Glewwe, 2000).

The mid-1980s saw a surge in new survey programmes and a large increase in the number of LMICs conducting household surveys. In 1982, CDC replaced the FP-MCH survey with the Reproductive Health Survey (RHS), and in 1985, they launched a version of the RHS tailored to women and men age 15 to 24 called the Young Adult Reproductive Health Survey (YARHS) (Morris, 2000). In 1984, USAID launched a follow-on programme to the WFS called the Demographic and Health Survey (DHS) which is still in operation today, expanding the WFS about fertility, family planning, and child mortality to also include questions on health and nutrition (Boerma and Sommerfelt, 1993). In 1985, the World Bank launched the Living Standards Measurement Survey (LSMS) also still in operation today with questions about household consumption, income, employment, housing conditions, and health (Grosh and Glewwe, 1995). By 1995, nearly all countries worldwide had had at least one national demographic, health, or economic survey (Figure 32).

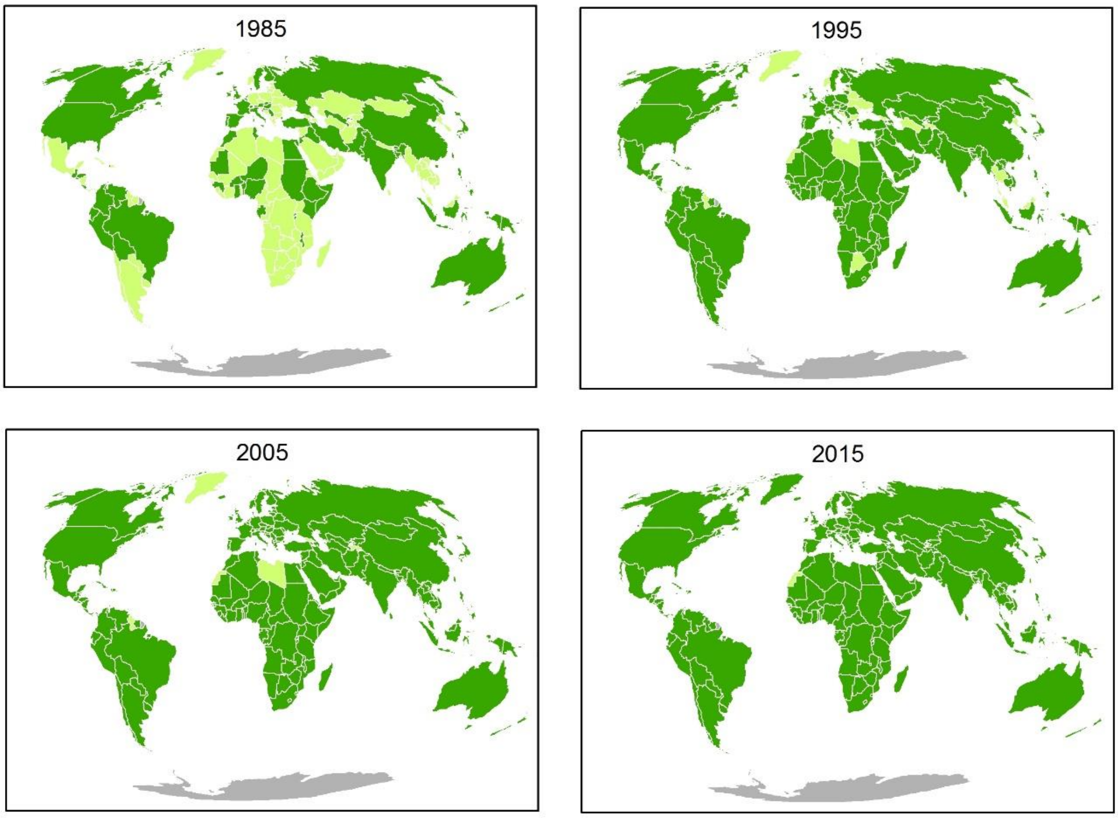


Figure 32. Coverage of at least one national household survey, by decade

Source: GHDx database (IHME, 2020). Dark green indicates coverage.

Still, large data gaps about health and wellbeing existed in the 1990s. Following the 1990 World Summit for Children, the UN Children's Emergency Fund (UNICEF) assessed available data on child health indicators across 99 countries and found that most survey data were too outdated to be useful (e.g., a DHS has been conducted more than five years earlier) or that existing surveys lacked important indicators (e.g., school enrolment, Vitamin A supplement, stunting) and/or did not cover children over the age of five (UNICEF, 2015). In fact, the only data widely available about child health in 1995 in LMICs was immunization coverage, mainly from EPI surveys (UNICEF, 2015). Thus, in 1995, UNICEF launched the Multiple Indicator Cluster Surveys (MICS) which covered nutrition and education indicators in children up to age 15, fertility and reproductive health of women, and household living conditions (UNICEF, 2015). Given substantial overlap in the target populations and content of the DHS and MICS, the two programmes have coordinated to ensure, for the most part, that their surveys are not conducted in the same countries, and today two-thirds of the questions in their questionnaires are identical or can be directly compared (Lisowska, 2016).

The 1990s saw a rise in biometric data collection in LMIC national household survey programmes. This is because as LMICs urbanized, the double burden of infectious and chronic diseases rose (Marshall, 2004; Bygbjerg, 2012). The WHO introduced the STEPS Adult Risk Factor Surveillance surveys in 2002 with methods to administer a questionnaire and take both physical and biochemical measurements from a sample of adults (Riley *et al.*, 2016). In the late 1990s, the DHS added several new modules to its standard questionnaire to respond to changing health needs, including blood sample collection to test for anaemia, malaria, HIV, and other key conditions, a domestic violence module, and calculation of a widely used assets-based wealth index (Fabic, Choi and Bird, 2012). Following the establishment of the Global Fund to Fight AIDS, Tuberculosis and Malaria in 2002, which has since provided billions of pounds in health funding to LMICs, shorter, topic-specific versions of the DHS have been implemented in select countries to monitor targeted indicators: the Malaria Indicator Survey (MIS) and AIDS Indicator Survey (AIS) (Fabic, Choi and Bird, 2012).

Today the DHS, MICS, and LSMS are the main sources of demographic, health, and economic data collected in LMICs. The EPI programme is still highly active, and in 2015, it underwent a major transformation and adopted probability-based sampling methods like those used in DHS, MICS, LSMS, and others (WHO, 2018). STEPS data are also still routinely collected today, though not in as many countries as DHS, MICS, and LSMS, and the STEPS datasets are rarely released publicly. Thus, the remainder of this chapter will focus on the DHS, MICS, and LSMS programmes.

The DHS, MICS, and LSMS programmes collectively cover more than 130 countries. Nearly 70 countries have participated in two of the three survey programmes, and 19 countries have participated in all three of the survey programmes (Development Initiatives, 2017) (Table 26). Survey sample sizes across all programmes are driven by the number of subnational areas (e.g., provinces, districts) and sub-populations (e.g., urban/rural) to be represented in the final results. All of these survey programmes adhere to the sampling methods and designs introduced in the 1970s and 1980s: two- or three-stage cluster sampling from small geographic areas – generally census enumeration areas from the last census – with household listing based on hand-drawn maps, and face-to-face interviews from a sample of selected households (Grosh and Munoz, 1996; ICF International, 2012a; UNICEF, 2013). Each of the survey programmes includes roughly a dozen modules with hundreds of questions, and each interview takes several hours to conduct. The number of households sampled in DHSs, MICSs, and LSMSs vary widely by country and year, and have generally increased over time with the need for further disaggregated national statistics. DHS surveys, for example, range from just over 2,000 households in the 1998-99 Cote D’Ivoire survey (INS and ORC Macro, 1999), to over 600,000 households in the 2015-16 India survey (IIPS and ICF International, 2017).

Table 26. Summary of DHS, MICS, and LSMS household survey coverage and sample designs

Survey Programme & Years	**Countries & Surveys** *by Dec 2018*	**Coverage & Strata**	**Sample Frame & Target Population**
DHS, 1985-present	86 countries 272 surveys (DHS Program, 2019)	National by province or district	Census, All women age 15-49
LSMS, 1985-present	37 countries 108 surveys (World Bank, 2018)	National by province or district	Census, Households
MICS, 1995-present	64 countries 293 surveys (UNICEF, 2019)	National by province or district, or province only	Census, One woman age 15-49

Household surveys in LMICs have not undergone the changes experienced in HICs toward mail, phone, or internet based surveys because a large portion of households in LMICs still lack a mailing address, phone, or the internet. Although access to mobile phones and the internet are increasing rapidly, huge disparities exist between and within LMICs. The International Telecommunications Union estimates that there are 75 mobile phone subscriptions per 100 people across Africa, with a range of 12 (Angola) to 162 (Senegal) subscriptions per 100 people (International Telecommunications Union, 2018). Subscription counts per population cannot be interpreted as percent coverage, as some people own multiple phone subscriptions, and many others have none. Those who do not have a mobile phone subscription are overwhelmingly the poorest, experience the worst health outcomes, and face the most barriers to accessing services (Jennings *et al.*, 2015). Thus, use of random digit dialling to mobile phones is not recommended for survey sampling in LMICs (Gibson *et al.*, 2017). However, use of mobile phones to follow participants and collect data in longitudinal surveys when a mobile phone is provided at no cost in an initial face-to-face interview can be effective. This approach has been tested with success by the World Bank for a household living conditions surveys in select countries with reliable mobile phone service, though it is worth noting that a traditional area-based sample frame is still used to identify an initial sample of households in these surveys (Dabalen *et al.*, 2016). Perhaps the greatest innovation in LMIC household surveys since they began is the use of modelled gridded population datasets as a sample frame in countries with outdated or inaccurate census data. This is the focus of the rest of this chapter.

5.2.3 Probability theory and survey design characteristics

Throughout the evolution of typical household surveys, important theories and survey design characteristics emerged. I briefly summarize key ideas and approaches that developed in the field before introducing the emerging field of gridded population sampling.

Probability sampling. A "good" sample is not biased, meaning that it represents a microcosm of the population, or universe, from which it was drawn. Probability sampling means that every unit in the population (e.g., household) has a known, non-zero probability of being sampled before the sample is drawn (Lohr, 2009). Random selection is a necessary condition for a probability sample because it prevents conscious bias. Non-randomised sampling, for example purposefully selecting units from the population, or sampling units that are convenient, might also have non-zero, known probabilities of selection, but the lack of randomisation can lead to bias.

Simple random sampling. A simple random sample is the most basic form of a randomized sample. In a simple random sample we make a list of all units in the population, and select a set number of units at random with or without replacement.

Stratification. Strata refer to non-overlapping groups that comprise the entire population (Lohr, 2009). Operationally, to stratify is to select an independent sample in groups within the population. Stratification is used for one or more of the following reasons.

First, to ensure representation of important groups in the sample (Lohr, 2009). For example, in most countries, urban and rural population have very different characteristics, risks, and outcomes. It is possible that, by chance, no households are sampled from one of the groups (i.e., urban or rural) in a given random sample of households. Stratification calibrates the sample by ensuring a specified number of units are sampled from each group, reducing the probability that any one draw will result in an unlucky, unrepresentative sample.

Second, stratification is used to achieve a set level of precision in the estimates for each group (Lohr, 2009). This is useful if we want to compare group characteristics in the population. In our example of urban and rural households, stratification on urban/rural could ensure that an estimate for urban households has similar precision as estimates for rural households.

A third reason to use stratification is feasibility of implementing the survey (Lohr, 2009). In national household surveys, it is often logistically more feasible and cost effective to have state/provincial statistical offices administer the survey, so administrative units would be used to stratify the national sample.

Finally, stratification can increase the statistical power of a sample if the units within each stratum are more similar to each other than to units in the rest of the population (Lohr, 2009). Statistical power is gained when the variance of indicators within each stratum is lower than the variance of indicators in the population overall.

Household surveys often stratify on urban/rural and by first or second administrative region for all of these reasons: to ensure coverage, increase precision for group comparisons, facilitate implementation, and to increase statistical power, especially in countries with major regional population differences.

Multi-stage cluster sampling. Every time that all units in a population are listed and then sampled, we call this a stage of sampling. Multi-stage sampling refers to surveys where more than one list is created and sampled, for example, a list and a sample of enumeration areas, followed by a list and a sample of households (see Figure 33). The first set of samples is referred to as the primary sampling unit (PSU), the second set of samples is called the secondary sampling unit (SSU), and so on. Multi-stage sampling is used for two reasons.

First, because a list, or sample frame, of the units that the survey team wishes to sample does not exist (Lohr, 2009). Until recently, many census agencies did not have a complete list of households with geo-coordinates or unique address locations by which to locate households for interviews (UNFPA, 2019), and even in the upcoming 2020 census round, many LMICs will not have these data in a format suitable for a household survey sample frame. Thus, the smallest administrative unit for which population counts are recorded by census – usually enumeration area (EA) – are used as the first-stage sample frame. Then a second list of households is developed for each sampled EA (i.e., cluster, PSU), and sampled. Even in countries with a full geo-located list of households, survey planners would likely not use it because it would be outdated by the time of survey fieldwork.

The second reason to use multi-stage cluster sampling is feasibility of implementing the survey (Lohr, 2009). A simple random sample of households in a province or country would simply cost too much money and time to justify, as interviewers would need to travel potentially long distances between each sampled household. Cluster sampling enables field teams to plan logistics in one village or neighbourhood for several days at a time.

More than two stages of sampling can be performed; in countries with highly aggregated census population counts, or countries with extremely long EA sample frames, survey teams may sample a higher-level administrative unit, such as district, first, then sample smaller areas, before sampling households (e.g., (MINSALUD and Profamilia, 2015)).

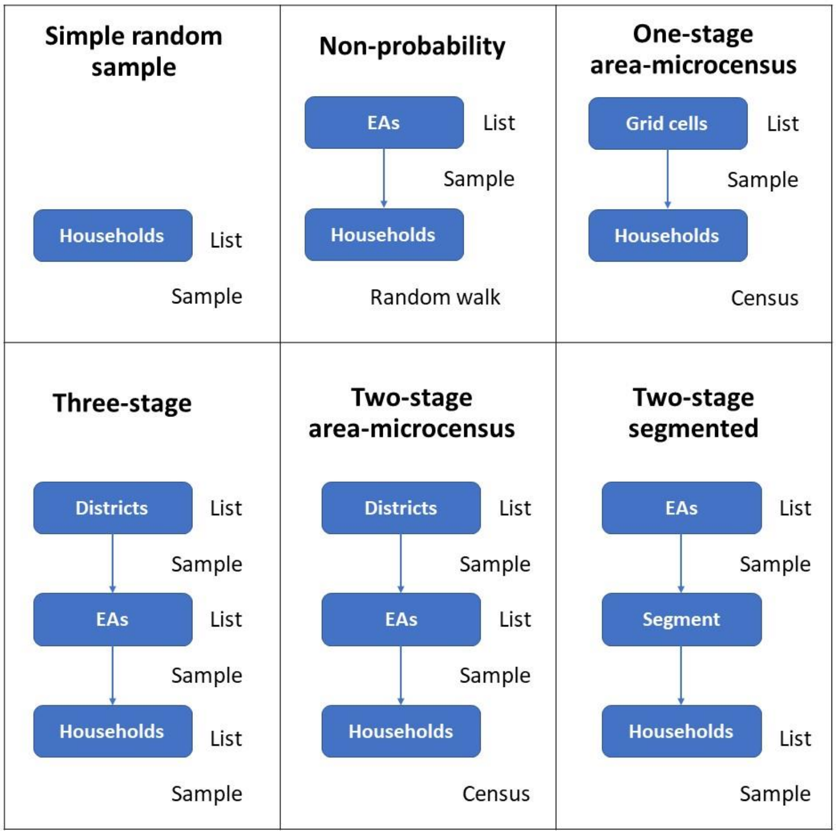


Figure 33. Depiction of select survey sampling concepts

Area-microcensus. This is not a term widely used by survey statisticians, but I use it throughout this thesis to refer to single-stage cluster surveys in which all households are sampled in a small area. The term “census” indicates that there is no further listing and sampling steps (Figure 33). I specify “area”-microcensus to differentiate the concept from other uses of the term “microcensus.” Census agencies use the term “microcensus” to both refer to a census of all residents in a small area (e.g., Alba, Muller and Schimpl-Neimanns, 1994), as well as a census of all residents in a sample of households (e.g., Meraner, Gumprecht and Kowarik, 2016).

Segmentation. Segmentation is a step performed in the field when a sampling area (i.e., cluster, PSU) is found to have far more people than expected, for example, due to a new housing development or emergence of a slum. In surveys such as DHS, MICS, and LSMS that perform full household listings before sampling households, there are usually not enough resources and time allocated to list more than 200 to 300 households per PSU. Thus the mapping-listing team divides the area into two or more approximately equal-sized segments, and randomly selects one segment to represent the PSU (ICF International, 2012a). As long as the segments have approximately equal population totals, probability of household selection can be calculated by recording the number of segments (because number of households in the listed segment, multiplied by the number of segments, equals approximately the number of households in the originally sampled EA). Segmentation is considered an additional stage of sampling if it is performed in every unit as part of the sample design (e.g., early MICS surveys (UNICEF, 2000); however, it is treated as a separate step when performed *ad-hoc* in large sampling units in the field (e.g., DHS (ICF International, 2012a)).

Oversampling. Oversampling means that the sample size is boosted in part of the population. In countries with majority rural populations, it is routine to oversample urban areas if there are not enough resources to stratify by both administrative unit and urban/rural areas. The purpose of oversampling, in this case, is to produce a sufficient sample size in the smaller important sub-group to generate precise estimates at the national level (ICF International, 2012a). In HICs, some health surveys oversample geographic regions where minority groups tend to live to boost the sample size of racial, ethnic, LGBT, and other sub-populations (Chen and Kalton, 2015; Anderssen and Malterud, 2017).

Although rarely practiced in human population surveys, another reason to oversample is to ensure spatial coverage of the sample. This type of oversampling is more common in environmental and animal population surveys (Kermorvant *et al.*, 2019), though could be useful in human population surveys to improve errors of small area estimates generated with household survey data (Dana R Thomson, Rhoda, *et al.*, 2020).

Design effect. Stratification, cluster sampling, oversampling, and other complex survey designs modify the precision of household survey estimates. The design effect quantifies the difference in precision in a given survey's estimates compared to a hypothetical simple random sample of the same size (Lohr, 2009). The design effect varies by indicator, depending on the variability and pattern of dispersion of that indicator in the population. Thus, survey implementers often report the design effects (DEFFs) or square root of the design effects (DEFTs) of key indicators with their survey results. Design effect can be interpreted as a factor by which to increase the sample size calculated for a simple random sample to achieve the target level of precision in a sample drawn with a complex design (e.g., 95% confidence level). The planners of household survey review past surveys which used a similar sample design in a similar context, and use reported DEFFs/DEFTs to calculate sample size requirements for key indicators.

Coverage error. An ideal sample frames includes all units of the target population (e.g., census EAs) such that the units are exhaustive, non-overlapping, and uniquely identifiable; however, perfect frames are rare (Kish, 1965; UN, 1982). Known problems in population sample frames include under-coverage and over-coverage, and can occur at each stage of sampling. Under-coverage means that units which are missing from the frame (e.g., EAs in a disputed territory), but might be supplemented from a different sample frame and treated as a separate stratum (Kish, 1965). In the case that two combined frames overlap (over-coverage), the overlapping units in one frame can be weighted by the proportion of non-overlapping area, essentially subtracting duplicate populations in the combined frame (Kish, 1965). Under-coverage can also occur when the auxiliary information (e.g., population size) associated with each sample frame units (e.g., census EAs) are differentially too small (or large) in some units due to an outdated sample frame (UN, 1982). In practice, this is sometimes addressed by conducting pre-survey field enumerations in units that are known to have experienced substantial population change since the sample frame was developed (e.g., EAs in and around cities) (UN, 1982).

Under- and over-coverage can also occur during survey implementation if areal unit boundaries are not identified accurately in the field, resulting, for example, in a unit not being fully enumerated, or a field enumeration extending incorrectly beyond the unit boundary (UN, 1982). Coverage errors also occur at the household- or individual-level because a household has recently moved, or an individual moved between households. These coverage errors can be minimised with field protocols and strict definitions of the household and its members (e.g., usual residence (*dejure*) versus presence at time of survey (*defacto*)) to minimise the chance that any one person or household can be counted more than once (UN, 1982). An error increasingly common in cities today is under-coverage of individuals or households living in atypical dwellings (e.g., shops) because data collectors were not provided a protocol to identify these households (Thomson *et al.*, 2021).

Non-response error. Non-response error occurs when an individual refuses to participate in a survey, is unavailable or unable to participate in the survey (e.g. not at home, unwell), or stops responding part-way through the survey (Kish, 1965). Usually survey protocols require several follow-up visits to households that were unavailable to minimise this type of non-response error. It is important to document, as best as possible, all eligible respondents and the specific reason for non-response (e.g., refusal, unavailable, incomplete survey) so that the effects of different types of non-response on sample results can be assessed (Kish, 1965). The number of respondents completing a survey divided by the number of eligible respondents is called the "response rate."

There might be systematic geographic or social patterns in non-response rates (e.g., urban residents are less likely to be home during the day and more sceptical about answering the door for strangers than rural residents). Geographic patterns in non-response are easier to assess because we often have geographic information about the non-responding households, and can compare the geographic distribution of non-responding households to those who responded. However, non-response that follows social patterns (e.g. household wealth status) is difficult to assess because we do not typically have social data about the non-responding households to compare with respondents.

Sample probability weights. If the probability of selection for each unit in the population is equal and non-response occurred completely at random, we call this a self-weighting sample. In practice, surveys are rarely, if ever, self-weighting. Most surveys use a combination of stratification, oversampling, and segmentation which results in some households having a greater probability of selection than others, and essentially all surveys will face some level of non-response. Furthermore, unequal population growth since the last census (e.g., due to different birth rates or migration) will mean that the number of households observed in the field at the time of survey differ from the counts made in the last census, and used to select PSUs.

In this case, we calculate and apply a weight for each unit (e.g., household) in the sample to make unbiased estimates about the population. The sample weight for each unit (i) is the reciprocal of the probability that the unit was selected (π_i) (Kish, 1965):

$$w_i = \frac{1}{\pi_i}$$

In household surveys, sample weights account for the selection probability of the stratum (if samples were not allocated proportionally to strata population totals), the selection probability of a PSU (usually an EA population count), the selection probability of a household within the PSU (based on a field listing of all households in that PSU), and PSU and household response rates. A household survey sample weight would thus be calculated as:

$$w_i = \frac{H_k}{n_k \times h_{jk}} \times \frac{M_{jk}}{m_{jk}} \times b_{jk} \times \frac{n_k}{n_{k*}} \times \frac{m_{jk}}{m_{jk*}}$$

Where:

H_k is number of households in stratum k according to the sample frame

h_{jk} is the number of households in PSU j according to the sample frame

n_k is the number of clusters sampled in stratum k

n_{k*} is the number of sampled clusters that were found and visited in stratum k

M_{jk} is the number of households enumerated in PSU j in stratum k during fieldwork

m_{jk} is the number of sampled households in PSU j in stratum k

m_{jk*} is the number of sample households that responded in PSU j in stratum k

b_{jk} is the number of equal-sized segments created in PSU j in stratum k

Non-probability sampling. Non-probability sampling means that the probability of being sampled is unknown for each unit in the population (Lohr, 2009).

Convenience / purposeful sampling. The most obvious non-probability sample is one in which randomisation is not applied. For example, by sampling the households of people that you already know, or households along major roads, neither would provide data that could be generalized to the population.

Random-walk and spin-the-pen. Random-walk and spin-the-pen are field methods designed to randomly sample households in a PSU. In a random-walk, the interview team starts at a main street or intersection, and follows a strict pattern of left and right turns, skipping a set number of buildings or dwellings between interviews (UNICEF, 2000). Similarly, spin-the-pen literally means that the field team spins a pen to randomise the direction of their walk after each interview (Grais, Rose and Guthmann, 2007). If implemented strictly, random-walk and spin-the-pen can result in a random selection of households in a given PSU. However, these methods result in a non-probability sample because the probability of household selection is unknown without the full enumeration of all households in the PSU. In practice, survey teams that use these methods often apply sample probability weights to adjust for stratification and oversampling in earlier stages of the sample.

Random-walk, spin-the-pen, and other types of randomized field-based household selection methods are harshly criticised for their susceptibility to conscious and unconscious bias by fieldworkers to avoid undesirable households, which can lead to systematic bias toward middle-class or accessible households (Grais, Rose and Guthmann, 2007; WHO, 2018).

Quota sampling. In quota sampling, the population is divided into groups and a target number of samples is set for each group before sampling (Lohr, 2009). In LMIC household surveys, quota sampling is often combined with random-walk or spin-the-pen methods; for example, in old EPI surveys, a quota of seven children between certain ages was set (Henderson and Sundaresan, 1982). The field protocol was to continue the random-walk/spin-the-pen method until the set number of households were identified and interviewed to meet the quota. Not only was this problematic because the sample probabilities of sampled households could not be known, the protocol also ignored eligible households that refused or were unavailable, biasing the survey toward the types of households who were at home at the time of interview and responded (WHO, 2018).

5.3 Gridded population sampling

The rest of this chapter is focused on gridded population surveys, defined as any sample drawn from a gridded population sample frame.

5.3.1 Literature review

To identify existing gridded population surveys, I conducted a literature review in Scopus using the terms: ("gridded" OR "landscan" OR "worldpop" OR "gpw" OR "ghs-pop" OR "hrsl" OR "wpe" OR "demobase") AND ("population" OR "household") AND "survey." This resulted in 65 potential articles as of May 2019. I screened all article abstracts and retained any that referred to sampling of human populations, resulting in 13 publications. I performed a full-text review of all screened articles and reports, and retained six that described a method, tool, or survey sampled from gridded population data. I additionally solicited colleagues for additional reports, websites, and articles describing a gridded population survey method or implementation, resulting in six additional resources (Figure 34).

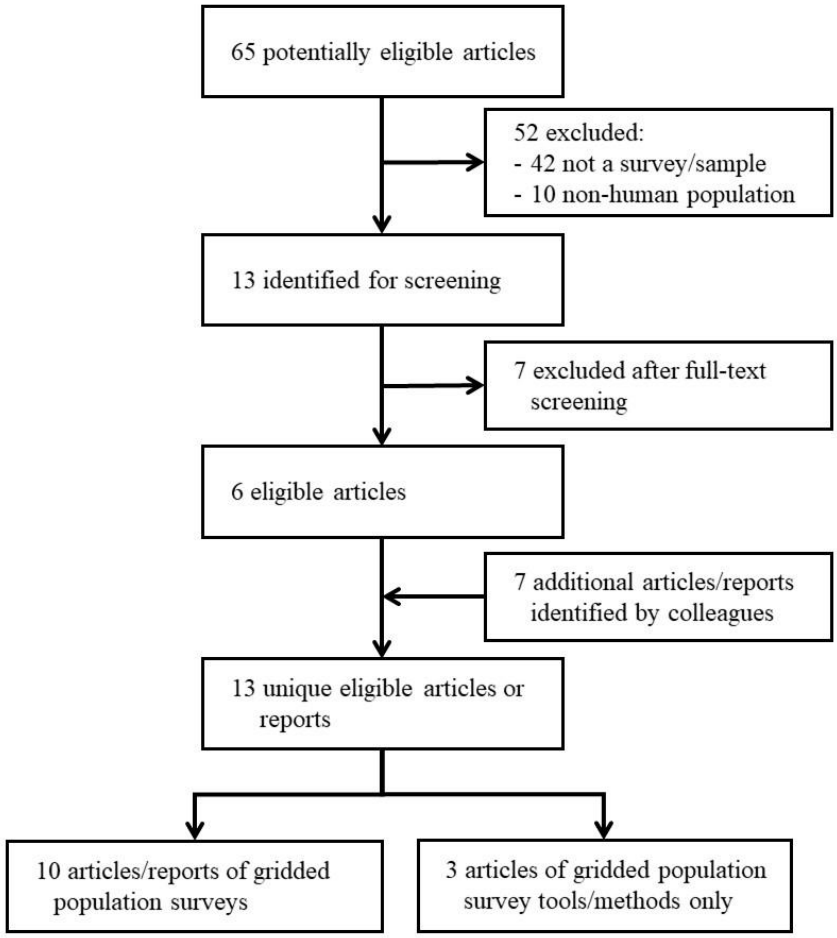


Figure 34. Gridded population survey literature review workflow and results

This literature review resulted in 13 articles and reports describing 37 gridded population surveys across 22 LMICs and 3 HICs. The surveys in LMICs spanned Asia (Bangladesh, India, Indonesia, Myanmar, Nepal, Thailand, Vietnam), the Middle East (Iraq), Africa (DR Congo, Ghana, Kenya, Ivory Coast, Mozambique, Nigeria, Rwanda, Somalia, Tanzania, Togo, Uganda), and Latin America (Brazil, Colombia, Guatemala) (Table 27). Three documents described tools or methods for selecting gridded population survey clusters (Muñoz and Langeraar, 2013; Thomson *et al.*, 2017; Chew *et al.*, 2018) (Table 27). Although gridded population surveys have been conducted in three HICs (Greece, Italy and Slovenia), they are not reported here.

Twenty-four of the 37 surveys had national coverage with 100 to 405 clusters each (Table 27). The first national gridded population survey was conducted in Iraq in 2011 by an academic team evaluating mortality rates in collaboration with Iraqi government officials (Galway *et al.*, 2012; Hagopian *et al.*, 2013). This survey followed a two-stage design from LandScan-Global 1 kilometre by 1 kilometre grid cells, with a random-walk to identify households in the field to minimize risks for fieldworkers (Galway *et al.*, 2012; Hagopian *et al.*, 2013). The firm RTI supported a commercial client to design and implement 12 national surveys and 1 sub-national survey across 11 countries in approximately 2014 and 2015 (year not reported) using a two-stage sample design based on LandScan-Global 1 kilometre by 1 kilometre grid cells, where an area-microcensus of eligible households was performed in the secondary sampling units (Cajka *et al.*, 2018). The company Gallup used gridded population sample frames in 10 LMIC and 3 HIC World Polls in 2017, 2018, and 2019 (with additional gridded population surveys planned for 2020) using WorldPop-RF 100 metre by 100 metre estimates for the first or second stage of sampling, and a random walk in the field to identify households (Gallup, 2017b).

Most sample frames in early gridded population surveys were derived from LandScan-Global 1 kilometre by 1 kilometre estimates, however most of the recent gridded population surveys derived sample frames from WorldPop 100 metre by 100 metre estimates (Table 27). Nineteen surveys performed an area-microcensus of households, 12 used a random-walk or spin-the-pen method to sample households in the field, and three subnational surveys performed a full household listing and sampling activity (Table 27).

A fundamental difference between census and gridded population data is that census EAs are each comprised of approximately the same number of people within varying sized areas, while grid cells are essentially uniform in area but vary widely in population. Household survey sampling methodologies have developed around the concept of small areal units each containing approximately the same number of people. Thus, a key challenge in gridded population sampling is forming areas of approximately equal population from grid cells. A number of approaches have been taken, often depending on whether the survey planner started with LandScan 1 kilometre by 1 kilometre grid cells or WorldPop 100 metre by 100 metre grid cells. Figure 35 summarises the tools used and approaches taken in gridded population surveys to derive sample frame units.

***Ad-hoc* GIS Approaches.** Many gridded population survey teams developed their own ad-hoc approaches to sampling using GIS software, such as ArcGIS.

- Galway et al. (2012) sampled 1 kilometre by 1 kilometre cells directly with PPS, then randomly selected one household in one building and performed a random walk. The building was selected by overlaying a mini 10 metre by 10 metre grid (estimated to be the average building footprint size), and randomly selecting mini grid cells until the team observed a building with satellite imagery (Galway *et al.*, 2012; Hagopian *et al.*, 2013).
- A similar approach was used by Gallup; the team aggregated 100 metre by 100 metre gridded population estimates to larger cells (depending on population density), then sampled the aggregated grid cells will PPS, before randomly selecting a building to start a random walk in the field (Gallup, 2017b).

- Thomson et al. (2012) converted 1 kilometre by 1 kilometre population estimates to corresponding numbers of random points per cell, selected points at random, manually delineated clusters (within cells) around each selected point over satellite imagery, and then performed an area-microcensus in the field.
- Muñoz and Langeraar (2013) proposed an approach, though it is unclear if a survey followed. In this approach, 1 kilometre by 1 kilometre cells were aggregated to 3 kilometre by 3 kilometre grid cells and sampled with PPS. Next, 1 kilometre by 1 kilometre grid cells were combined within selected 3 kilometre cells to achieve a minimum population, and then sampled with PPS. The sampled 1 kilometre (or larger) areas were manually delineated into segments of approximately 100 households each over satellite imagery, and one segment was randomly selected. Finally, households were listed via a field mapping-listing activity, and a sample of households was selected.
- Sollom et al. (2011) joined 1 kilometre by 1 kilometre gridded population estimates to rural village point locations and sampled those points with PPS, and then used spin-the-pen to sample households in the field.
- Qader et al. (2019) used gridded population estimates to update census EA counts in urban areas where EA boundaries were available, and used a quadtree method to create different sized grid cells with similar population totals in rural areas. The combined frame was sampled with PPS before manually segmenting over satellite imagery and randomly selecting one household per segment.

The tools created for gridded population sampling included the GridSample R package (Thomson *et al.*, 2016, 2017), Geo-sampling Tool (Cajka *et al.*, 2018; Chew *et al.*, 2018), and GridSample.org (Flowminder Foundation, 2019a), which is based on the GridSample2.0 python algorithm (Flowminder Foundation, 2019b) (Table 28).

Table 27. Summary of gridded population survey coverage and sample designs identified in the literature review

Country & Year (if reported)	Design: Coverage, Strata, Stages	Cluster & Household Sample Size	Gridded population dataset	Target Population, Main topic(s)
DR Congo 2010 [1]	Idjwi Island, none, one-stage area-microcensus	50 clusters, 2078 HHs	2001 LandScan-Global	All women age 18-50, Maternal and child health
Myanmar 2010 [2]	Chin state, urban/rural, two-stage (spin-the-pen)	90 clusters, 720 HHs	2005 LandScan-Global (rural only)	Household head age 18+, Health, human rights
Iraq 2011 [3]	National, governorates, two-stage (random-walk)	100 clusters, 1960 HHs	2008 LandScan-Global	Household head age 18+, Mortality
Bangladesh [4]	National, division x urbanicity, two-stage area-microcensus	148 clusters, 3296 HHs	2012-2016 LandScan-Global	Adult age 18+, topics not reported
Brazil [4]	National, region x poverty, two-stage area-microcensus	149 clusters, 3652 HHs		
Colombia [4]	National, region x poverty, two-stage area-microcensus	152 clusters, 2706 HHs		
Colombia [4]	National, region x poverty, two-stage area-microcensus	152 clusters, 3037 HHs		
Ghana [4]	National, region x poverty x urbanicity, two-stage area-microcensus	151 clusters, 3113 HHs		
Guatemala [4]	National, department x urbanicity, two-stage area-microcensus	211 clusters, 3057 HHs		
India [4]	Three states, district x urbanicity, two-stage area-microcensus	467 clusters, 10,824 HHs		
Kenya [4]	National, province x poverty, two-stage area-microcensus	143 clusters, 3364 HHs		
Nigeria [4]	National, region x poverty, two-stage area-microcensus	147 clusters, 3042 HHs		
Rwanda [4]	National, province x poverty, two-stage area-microcensus	150 clusters, 3096 HHs		
Thailand [4]	National, region x poverty, two-stage area-microcensus	150 clusters, 3136 HHs		
Thailand [4]	National, region x poverty, two-stage area-microcensus	150 clusters, 3275 HHs		
Uganda [4]	National, region, two-stage area-microcensus	146 clusters, 3075 HHs		
Nepal 2015 [5]	Kathmandu, none, two-stage	90 clusters, 1,310 HHs (planned)	2014 WorldPop-RF	Woman age 18+, Maternal and child health

Mozambique 2017 (6)	Six districts, district, one-stage area-microcensus	234 clusters, 4998 HHs	2017 WorldPop-RF	Caregiver of child age 12-18, Child health
DR Congo 2017 (6)	Kinshasa, communes, one-stage two-stage area-microcensus	210 clusters, 1,850 HHs	Bespoke	Household head, Food insecurity
Somalia 2017 (7,8)	National, region x urbanicity, two-stage	405 clusters, 6,284 HHs	Modified 2015 WorldPop-LC	Household head, Economic
Nepal 2017 (6,9,10)	Kathmandu, none, one-stage area-microcensus	30 clusters, 600 HHs	2017 WorldPop-RF	Adult age 18+, Economic, non-communicable disease
Nepal 2017 (6,9,10)	Kathmandu, none, two-stage	30 clusters, 600 HHs		
Bangladesh 2018 (9,10)	Two communities, community, one-stage area-microcensus	20 clusters, 400 HHs	2020 WorldPop-RF	
Vietnam 2018 (9,10)	Long Bien District, none, one-stage area-microcensus	20 clusters, 400 HHs		
Colombia 2017 (11)	National, region x urbanicity, two-stage (random walk)	125 clusters, 1000 HHs	2015 WorldPop-RF	Adult age 15+, Topics not reported
Tanzania 2017 (11) *	National, region x urbanicity, three-stage (random walk)	400 clusters, 4000 HHs	2015 WorldPop-RF	
Uganda 2018 (11)	National, region x urbanicity, two-stage (random walk)	200 clusters, 2000 HHs	2020 WorldPop-RF	
Nigeria 2018 (11)	National, region x urbanicity, two-stage (random walk)	300 clusters, 3000 HHs	2020 WorldPop-RF	
Indonesia 2018 (11)	National, region x urbanicity, two-stage (random walk)	400 clusters, 4000 HHs	2015 WorldPop-RF	
Colombia 2018 (11)	National, region x urbanicity, two-stage (random walk)	400 clusters, 4000 HHs	2020 WorldPop-RF	
Kenya 2018 (11)	National, region x urbanicity, two-stage (random walk)	200 clusters, 2000 HHs	2015 WorldPop-RF	
Ghana 2019 (11)	National, region x urbanicity, two-stage (random walk)	100 clusters, 1000 HHs	2020 WorldPop-RF	
Togo 2019 (11)	National, region x urbanicity, two-stage (random walk)	100 clusters, 1000 HHs	2020 WorldPop-RF	
Ivory Coast 2019 (11)	National, region x urbanicity, two-stage (random walk)	100 clusters, 1000 HHs	2020 WorldPop-RF	

1. (Thomson *et al.*, 2012)
2. (Sollom *et al.*, 2011)
3. (Galway *et al.*, 2012; Hagopian *et al.*, 2013)
4. (Cajka *et al.*, 2018)
5. (Elsey *et al.*, 2016)
6. (WFP-VAM, 2018; GridSample, 2019)
7. (Pape and Wollburg, 2019)
8. (Qader *et al.*, 2019)
9. (Elsey *et al.*, 2018)
10. (Dana R. Thomson *et al.*, 2020)
11. (Gallup, 2017b)

* Gridded population sample frame used in second stage of sampling

Figure 35. Summary of methods used to create gridded population survey sample frame units

Adapted with permission from (Thomson et al., 2017)

Table 28. Comparison of sampling tools used in gridded population surveys

Feature	GridSample R	Geo-sampling	*Ad-hoc* GIS	GridSample2.0	GridSample.org
Public	Yes	No	Yes	Yes	Yes
Free	Yes	No	Some	Yes	Yes
Skill level required	Advanced	Advanced	Advanced	Advanced	Basic
User selects the sample	Yes	No	Yes	Yes	Yes
Gridded pop	Any	LandScan-Global	Any	Any	WorldPop-Global
Preloaded/ provided data	No	Yes	Some	No	Yes
Pre-forms clusters	No	Yes	Some	Yes	Yes

GridSample R package. I released the GridSample R package with colleagues in 2016 (Thomson *et al.*, 2016, 2017) which has been used in at least six sub-national surveys (Elsey *et al.*, 2016, 2018; GridSample, 2019). The algorithm treats the gridded population dataset as the sample frame and selects grid cells with PPS allowing for stratification, oversampling in urban/rural domains, and spatial oversampling (Thomson *et al.*, 2017). The GridSample R package runs on a personal computer and is limited by the computer's memory. All datasets must be pre-processed and specified by the user, allowing use of any gridded population dataset but also requiring GIS and/or R programming skills. The algorithm enables optional "growth" of clusters to a minimum population size or maximum area by randomly adding neighbouring cells after selection of "seed" cells with PPS. While this process results in clusters with roughly consistent population sizes for improved fieldwork, the population counts in the "grown" clusters do not reflect the population counts used for sample selection, and thus may skew sample weights (Thomson *et al.*, 2017). The output is a shapefile of cluster boundaries, with attributes of estimated population counts. Thomson, et al. (2019) used this tool to sample grid cells directly in city-level surveys in Bangladesh, Nepal, and Vietnam, whereas Elsey, et al. (2016), World Food Programme VAM unit (2018), World Vision International (GridSample, 2019), and Thomson, et al. (2019) "grew" clusters to contain 100 to 200 households each for sub-national surveys in Nepal, DR Congo and Mozambique.

Geo-sampling Tool. The Geo-sampling survey tool was created by RTI and used in 13 national and sub-national surveys (Cajka *et al.*, 2018). It is designed for use with 1 kilometre by 1 kilometre grid cells, and supports a multi-stage stratified sampling approach. After administrative units are sampled with PPS, grid cells are sampled with PPS. To improve fieldwork, the team excludes 1 kilometre cells with fewer than 250 estimated people, potentially biasing the sample toward higher-density populations. The sampled grid cells are then partitioned into 150 metre, 100 metre, or 50 metre grid cells depending on population density. Next, a deep-learning residential scene classification model is used with satellite imagery layers to exclude smaller cells without settlement, and disaggregate the 1 kilometre by 1 kilometre grid cell population estimates to the remaining smaller cells. Finally, three of the smaller cells are selected at random and an area-microcensus of households is conducted in the field (Chew *et al.*, 2018). The Geo-sampling Tool is an in-house RTI product, and clients are provided with a shapefile of the final cluster boundaries and population estimates.

GridSample.org and GridSample2.0. GridSample.org is a free web-based tool released in late 2019 that runs the open-source GridSample2.0 algorithm that I developed at Flowminder Foundation (Flowminder Foundation, 2019a, 2019b). GridSample.org provides a point-and-click interface, preloaded datasets, and guidance to enter parameters and select clusters for a gridded population survey. The website leverages gridEZ, a publicly-available algorithm developed by Dr. Claire Dooley, to group cells into gridded EA-like clusters with similar population totals before sampling (Dooley, 2019). Preloaded datasets include WorldPop-Global 100 metre by 100 metre gridded population estimates (WorldPop, 2019), GADM administrative boundaries (GADM, 2015), and GHS-SMOD urban/rural boundaries (European Commission, 2017).

GHS-SMOD is a global model of human settlement types produced by the European Commission Global Human Settlement Layer Project based on the project's gridded population dataset (GHS-POP) and its model of built areas (GHS-BUILT). GHS-SMOD is formatted as 1 kilometre by 1 kilometre cells classified as either high-dense urban, low-dense urban, rural, or unsettled/remote (European Commission, 2017). In GridSample.org, all surveys are implicitly stratified by level of urbanicity based on these four GHS-SMOD classes. GridSample.org also supports stratification and spatial oversampling; and custom coverage, strata, or sample frame boundaries can be uploaded by users as a shapefile. The website is designed for low-bandwidth settings, running sample selection remotely on a super-computer. The user is emailed a shapefile of cluster boundaries, excel table with population estimates to calculate sample weights, and a PDF report of survey parameters.

Field tools. A range of simple-to-advanced tools have been used to implement gridded population surveys. Lower-tech field tools included use of paper maps displaying cluster boundaries over satellite imagery produced in Google Earth, and paper listing forms and questionnaires (Galway *et al.*, 2012; Thomson *et al.*, 2012; GridSample, 2019). Higher-tech field tools included tablet-based applications for navigation (Cajka *et al.*, 2018; Thomson, Bhattarai, *et al.*, 2020), paper field maps designed in GIS (Galway *et al.*, 2012; Elsey *et al.*, 2016; GridSample, 2019; Qader *et al.*, 2019; Dana R. Thomson *et al.*, 2020), and tablet-based household listing and/or questionnaires (Elsey *et al.*, 2016; Gallup, 2017b; Cajka *et al.*, 2018; GridSample, 2019; Dana R. Thomson *et al.*, 2020).

Satellite imagery was essential to all gridded populations surveys to manually segment along roads, rivers, and other features (Thomson *et al.*, 2012; Muñoz and Langeraar, 2013; Elsey *et al.*, 2018), and as a field map base layer for navigation and to ensure that all eligible dwellings were included (Galway *et al.*, 2012; Thomson *et al.*, 2012; Cajka *et al.*, 2018; GridSample, 2019; Qader *et al.*, 2019). In some surveys, satellite imagery was used to digitize building footprints and roads in OpenStreetMap which were then displayed as a field map base layer (Elsey *et al.*, 2016, 2018). The RTI surveys used satellite imagery and machine learning to exclude unsettled areas from the sample and further disaggregate population estimates (Chew *et al.*, 2018).

5.4 Accuracy of typical versus gridded population sampling

Based on findings in Chapter 4 which identified sharp difference in cell-level accuracy by urban/rural location, this analysis of gridded population sampling accuracy is constrained to urban Khomas. A further reason for this decision is that relatively few households in Khomas are rural (5% of the simulated “true” population), which means an enormous oversample would be needed in rural areas to make estimates for the rural population within Khomas. The 2013 Namibia DHS, for instance, drew just 3 rural clusters and 50 urban clusters within Khomas (MoHSS and ICF International, 2014).

I chose to sample 60 PSUs, and then sample 20 households per PSU, for a target sample of 1,200 households overall. This is consistent with sample designs used in recent demographic surveys in Namibia including the 2013 Demographic and Health Survey (53 PSUs, 20 households per PSU) and 2016 Intercensal Demographic Survey (69 PSUs, 20 households per PSU) (Table 29). Recent economic surveys tended to have more PSUs and fewer households per PSU, with approximately 1,100 households sampled from Khomas overall (Table 29).

Table 29. Sample sizes in Khomas, Namibia in recent household surveys

Recent surveys in Namibia	PSUs in Khomas	Households per PSU	Khomas sample size
Demographic and Health Survey – 2013 [1]	53	20	1,060
Household and Income Expenditure Survey – 2015/16 [2]	96	12	1,152
Intercensal Demographic Survey – 2016 [3]	69	20	1,380
Labour Force Survey – 2018 [5]	63	18	1,134

1. (MoHSS and ICF International, 2014)
2. (NSA-NPC, 2016)
3. (NSA, 2016)
4. (NSA, 2018)

5.4.1 Methods

The methods used in this analysis are summarised in Figure 36 and Figure 37. In the first step, I constrained all datasets to urban Khomas (Chapter 3). In this analysis, “urban” covered all EAs classified as urban in the 2011 Namibia census plus areas of expansion around settlements, as I observed in satellite imagery. The analysis was conducted in R 3.5.2 using the EA shapefiles that I modified during the simulation in Chapter 3.

To create PSUs with approximately equal populations from the gridded population estimates generated in Chapter 4, I used the gridEZ R algorithm version 1 released in 2019 (Dooley, 2019). The gridEZ algorithm produces gridded enumeration zones from a gridded population dataset such that each unit has a target population or a maximum area. I provided input to Dr. Claire Dooley, the algorithm creator, during its development; specifically, to define three pre-set gridEZ unit sizes (small, medium, and large) that can support several common household survey designs. From GitHub where the gridEZ algorithm was released, here are the general steps that the gridEZ algorithm follows (Dooley, 2019):

- Within strata (constituency boundaries intersected with GHS-SMOD “high dense urban” boundaries), create rectangular “clumps” defined by the gridEZ maximum area, or 5 kilometres by 5 kilometres, whichever is smaller
- Divide each clump into “blocks” based on the clump population vs. the target population
- Any block that has fewer than the target population is combined with the contiguous neighbouring block with the lowest population
- Any block with twice the target population is halved

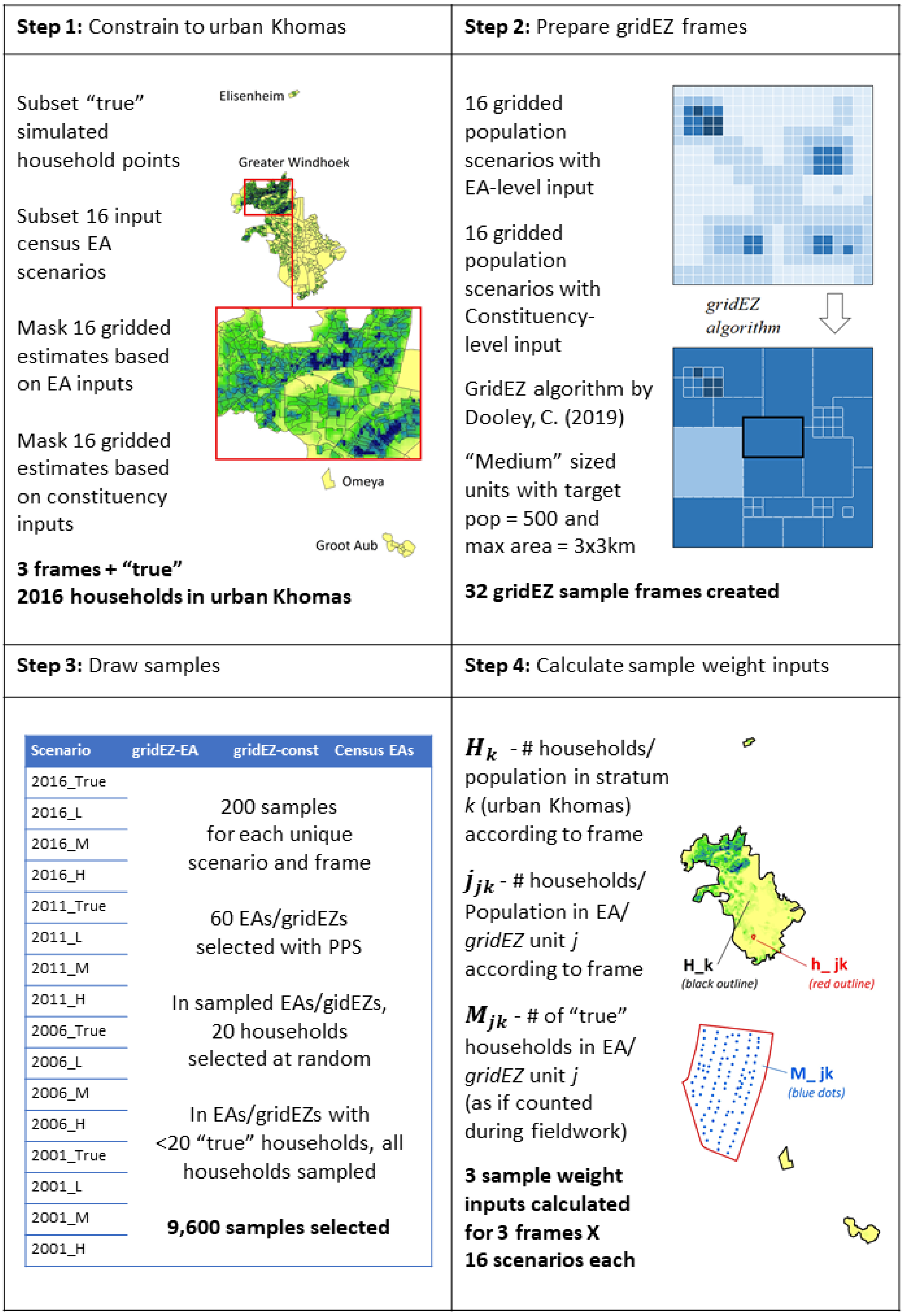


Figure 36. Steps 1 to 4 used to analyse accuracy of gridded population surveys

I used medium pre-defined gridEZ units for this analysis, each of which had a target of 500 people (approximately 135 households in urban Khomas) and a maximum area of 3 kilometres by 3 kilometres. In the gridEZ algorithm, I used constituency boundaries and GHS-SMOD “high dense urban” areas, clipped to urban Khomas boundaries to define the strata in which “clumps”, “blocks”, and gridEZ units were formed. I produced gridEZ units for each of the 16 gridded population datasets generated from EA-level input data (hereafter called gridEZ-EA), as well as each of the 16 gridded datasets generated from constituency-level input data (hereafter called gridEZ-constituency). This resulted in sampling units of similar population size across the census and both gridEZ sample frames.

The census and both gridEZ sample frames for the 2016 “true” population are summarized in Table 30. There were 766 units in the census frame, 786 in the gridEZ-EA frame, and 552 in the gridEZ-constituency frame (Table 30). The gridEZ-constituency frame had fewer units that met the 500 target population because the underlying urban gridded populations were underestimated due to misallocation of population to low-dense and unsettled cells outside of urban Khomas. GridEZ units were generated in R 3.5.2 using the code and instructions provided on Dr. Dooley’s GitHub page (Dooley, 2019).

Table 30. Comparison of three sample frames in urban Khomas based on the “true” 2016 simulated population

Characteristic	**Census EAs**	**GridEZ units** EA-level inputs	**GridEZ units** constituency-level inputs
Number of sample units in frame	766	787	590
Mean frame unit population (range)	455 (37 - 7270)	444 (1 - 1550)	437 (1 - 1259)
Mean frame unit area (range)	317 x 317 m (110x110m - 1943x1943m)	353 x 353 m (89x89m - 2051x2051m)	428 x 428 m (89x89m – 2051x2051m)

With the sample frames prepared, the third step was to draw samples. I drew 200 samples for each of the sample frames (3) and outdated-inaccurate scenarios (16) resulting in 9,600 samples. The sample design was, again, 60 PSUs, and 20 households per PSU, for a target of 1,200 households per sample. PSUs were drawn with probability proportional to size (PPS) from EA/gridEZ sample frames, while households were drawn at random from the “true” 2016 household point locations. This represented the use of an outdated-inaccurate sample frame that may not match the actual population identified during survey fieldwork. In cases where a PSU had fewer than 20 "true" households, all households were sampled. In practice, gridded population survey practitioners often draw extra back-up clusters, randomly select the target *n* main clusters from the draw, then drop and replace any main cluster with a back-up cluster if no habitable buildings are identified upon visual inspection of satellite imagery and/or during a field visit (Dana R Thomson, Rhoda, *et al.*, 2020; Thomson *et al.*, 2021). I implemented this approach by drawing 72 clusters (60 main + 20% back-up), dropped all clusters with no "true" households, and randomly selected 60 clusters from those that remained. I used R 3.6.0 for this step, drawing PPS samples with the *samplingbook* package `pps.sampling` function (Kauermann and Küchenhoff, 2011). I used the *spatialEco* package `points.in.poly` function to join "true" households to PSUs, and the base R `sample` function to draw simple random samples of households.

In the fourth step, I generated three of the inputs needed to calculate household probability weights: H_k, the number of households in urban Khomas according to the sample frame (or P_k, the gridded population estimate); h_{jk}, the number of households in the respective PSU according to the sample frame (or p_{jk}, the number of households in the gridEZ PSU); and M_{jk} the "true" number of households in the respective PSU from the simulated 2016 point file (Figure 36). R 3.6.0 was used for this step.

In step five, I calculated "true" household indicator values from the 92,844 simulated households in urban Khomas (Figure 37). I selected four socioeconomic household-level indicators for this analysis, three of which were described, evaluated and deemed to have realistic distributions in Chapter 3. These were percent of households with unimproved toilet, percent of households with crowding, and percent of households with a non-durable floor. Percent of households with unimproved water, which was also evaluated in Chapter 3, was excluded because none of the simulated households in urban Khomas had unimproved water. To replace this indicator, I selected an additional socioeconomic indicator that was simulated, but not evaluated: percent of households cooking with solid fuel such as charcoal or wood. This step was performed in R 3.6.0.

Step six is where I calculated sample probability weights and estimated indicators in each of the 9,600 samples. To do this, I calculated the last element of the sample probability weight: m_{jk}, the number of households sampled in the respective PSU. Sample weights for the i^{th} household in PSU j and stratum k followed the formula detailed in section 5.2.3 in census samples. Household sample weights in gridEZ units used P_k and p_{jk} population in place of H_k and h_{jk} households as shown below. Population estimates (P_k and p_{jk}) approximately equals households (H_k and h_{jk}) if we divide the population estimates by the average household size. Because average household size is in the numerator and the denominator of the household probability weights equation, it cancels out, allowing estimated population values to be used directly. I assumed no segmentation in sample frame units, and 100% household response:

$$w_i = \frac{P_k}{n_k \times p_{jk}} \times \frac{M_{jk}}{m_{jk}}$$

Where:

P_k is the estimated gridded population in stratum k

p_{jk} is the estimated gridded population in gridEZ unit j in stratum k

n_k is the number of gridEZ units sampled in stratum k

M_{jk} is the number of household point locations in gridEZ unit j in stratum k

m_{jk} is the number of sampled households in gridEZ unit j in stratum k

Following common survey practice, I normalised the household weights around one (ICF International, 2012a; UNICEF, 2013). Normalising weights reduces the effect of extremely small or large weights, which can occur in gridded population sampling because we find that a PSU derived from gridded data (and which has local spatial inaccuracy) has only a few households in the field, so we sample all or most of them. While raw sample probability weights sum to the total population in the sample frame, normalised sample weights sum to the total sample size. I normalised each weight for household i as follows:

$$w_{i.norm} = w_i \times \frac{\sum(m_{jk})}{\sum(w_i \times m_{jk})}$$

Where:

m_{jk} is the number of (responded) households in EA/gridEZ j in stratum k

w_i is the raw household probability weight

Given the focus of this thesis on representation of the urban poorest in household surveys, and the greater rates of missingness among simulated slum households, I calculated indicators by slum and non-slum in step seven (Figure 37). This represents a typical household survey in which slum locations are is not known in the sample frame, and thus can only be measured and analysed after sample selection. However, unlike a typical survey, I was able to calculate "true" slum and non-slum indicator values from the simulated population (*n* slum = 35,001; *n* non-slum = 57,843). For this analysis, I used slum household status of "true" 2016 household point locations created manually for each household in Chapter 3.

Maps are presented of select census EA and gridEZ sample frames to compare differences across scenarios of outdated and inaccuracy censuses. Comparisons are also made of the percent of slum/non-slum households in each sample versus the “true” percent of slum households, disaggregated by sample frame and outdated-inaccurate scenario in order to understand how the lack of within-urban stratification impacts disaggregated indicator estimates for slum and non-slum populations.

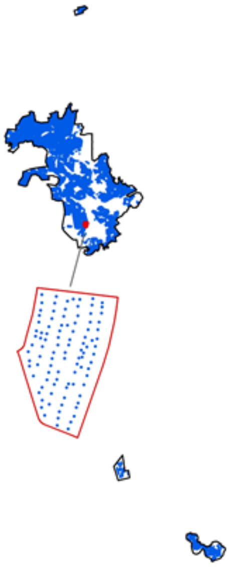
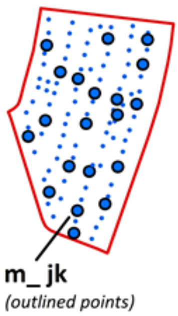
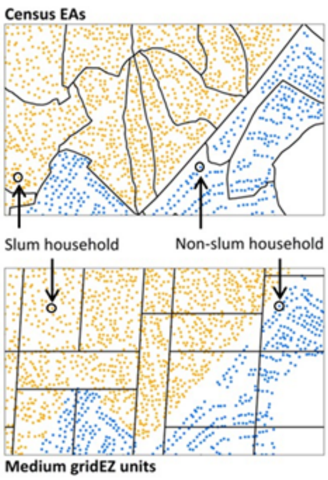
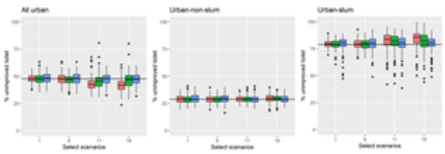
Step 5: Calculate "true" indicators	**Step 6:** Calculate weights & indicators
Calculate indicators for **2016 "true" population** in urban Khomas N = 92,844 Household indicators: • Unimproved toilet • Non-durable structure • Overcrowding • Solid fuel for cooking 	Join H_k, h_{jk}, and M_{jk} to each sampled household in 9,600 samples Calculate m_{jk} for each EA/gridEZ unit Assume no segmentation and 100% response rate  Calculate **household sampling weights**: $w_{jk} = \frac{H_k\ (or\ P_k)}{60\ \times h_{jk}\ (or\ p_{jk})} \times \frac{M_{jk}}{m_{jk}}$ Calculate **weighted indicators**, and 95% confidence interval accounting for clustering in each of the 9,600 samples
Step 7: Calculate slum/non-slum indicators	**Step 8:** Summary statistics
Household **slum / non-slum status** from "true" population used to classify sampled households 	**Box plots** for 9,600 samples: • By 3 frames • By 16 outdated-inaccurate scenarios • By slum / non-slum status  **Summary statistics for each set of 200 samples** in the 48 frame X scenario combinations, comparing estimated and "true" values **Sub-analysis** on slum/non-slum households Summary statistics: • RMSE • Bias

Figure 37. Steps 5 to 8 used to analyse accuracy of gridded population surveys

In a final step (step eight), I created boxplots and calculated summary statistics for each of the three sample frame and 16 outdated-inaccurate scenarios, as well as in slum and non-slum households (Figure 37). All indicators were summarized as weighted percentages, with 95% confidence intervals that accounted for clustering of households in PSUs. Indicator estimates were calculated in R 5.3.2 on the University of Southampton Iridis 5 high performance computing cluster using the *survey* package `svyciprop` function (Lumley, 2004). Slum and non-slum estimates were generated with the same package by adding a sub-population (*subset*) argument to the survey design specifications.

Given that many non-slum households had zero values for certain indictors, a likelihood-based approach to confidence interval estimation (e.g., Rao and Scott, 1987) was not ideal because it assumes an approximately normal distribution of the estimated proportion, and this assumption breaks down when the estimated proportion is close to zero (Korn and Graubard, 1998). Instead, I used the Korn and Graubard approach to estimate confidence intervals for proportions with small counts in complex survey data, specified in `svyciprop` as "beta." This approach uses a logit transformation and Poisson distribution, substituting *n* sampled observations for the effective sample size and its degrees-of-freedom when the estimated proportion is zero (Korn and Graubard, 1998).

Boxplots are presented to show the distribution of weighted indicators across the 100 samples for each of the 48 frame-scenario combinations, as well as in slum/non-slum sub-populations. Each plot displays these distributions in reference to the "true" indicator value marked with a horizontal line.

Two accuracy statistics were also calculated for each of the 48 frame-scenario combinations and slum/non-slum sub-populations to understand errors in gridded population samples, and to compare accuracy in census and gridEZ frames under realistic scenarios of outdated-inaccurate censuses. In accuracy statistics, each of the 200 sample indicator estimates were compared to the "true" indicator value, and then summarized. Root mean square error (RMSE), detailed in section 4.2.1.5, reflects the degree of accuracy (over- or under-estimation error) in the estimated indicators for a particular sample frame. Bias, detailed in section 4.4.1, indicates the direction and average error in a particular sample frame; whether it consistently leads to an over- or under-estimate.

5.4.2 Results and discussion

Visual inspection of the sample frames against the "true" 2016 household point locations revealed key differences between census-based sampling and gridded population sampling (Figure 38). The 2011 census EAs were drawn along locally meaningful neighbourhood boundaries which generally divided slum and non-slum areas. The simulated population thus largely segregated slum/non-slum households along EA boundaries. As expected, the distribution of slum/non-slum households in gridEZ units were more heterogeneous than in census EAs in urban Khomas (first column of Figure 38). GridEZ units were less spatially sensitive to missing populations than census EAs, even when highly outdated-inaccurate census counts were used as input to the gridded population model (second and third column of Figure 38). This is because the gridEZ algorithm groups cells across EA boundaries into units which are constrained only by higher-level administrative and GHS-SMOD “high-dense urban” boundaries (section 5.4.1).

When constituency-level inputs were used in the gridded population model, the most densely populated cells in Windhoek were substantially underestimated because population was misallocated to low-density or unsettled cells (see section 4.5.1). In scenarios of high census outdatedness-inaccuracy, this inadvertently gave excluded populations a greater probability of selection than they would have had if the census data are sampled directly, though it also increased the likelihood of sampling cells with no or few actual households.

GridEZ units derived from outdated, inaccurate census data had larger areas than gridEZ units derived from recent, accurate census data. The larger units were created by the gridEZ algorithm to maintain the target of 500 people per unit based on lower estimated population totals (second and third columns of Figure 38). With larger areas, these gridEZ units were more heterogeneous in terms of slum/non-slum households than gridEZ units derived from more accurate census data, or the census data itself.

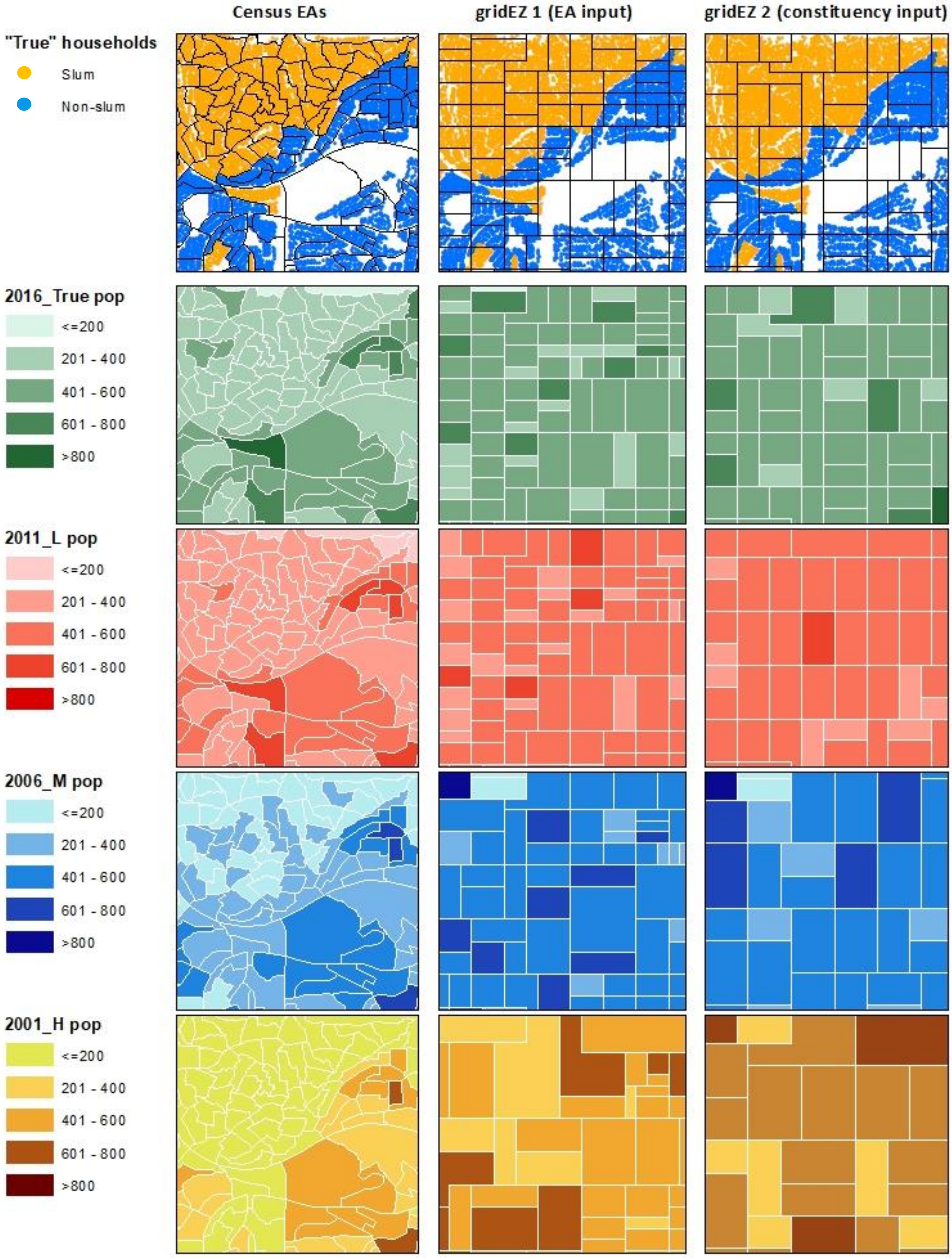


Figure 38. Visual comparison of the "true" 2016 household point distribution and three sample frames, by simulated census scenario in a section of Windhoek, Namibia

Figure 39 shows the distribution of slum/non-slum households across all samples, by sample frame and census scenario. In the best-case-scenario, samples drawn from 2016 frames with no population missingness (top-left plot in Figure 39), all estimated the true percent slum population on average, though gridEZ-constituency samples did so with more variability due to inaccuracies in the sample frame. Underestimation of the slum population increased in the census EA frame, and to a lesser extent the gridEZ-EA frames, as the census was increasingly outdated or inaccurate, with gridEZ-constituency frames consistently estimating the true slum population regardless the age or inaccuracy of the input population data (Figure 39). This strongly underscores an issue raised by urban health experts who argue that census-based household surveys (which overwhelmingly are based on several-year-old sample frames) and stratified by only urban/rural areas tend to produce overly cheery estimates of health and social outcomes in urban population (Lilford *et al.*, 2017).

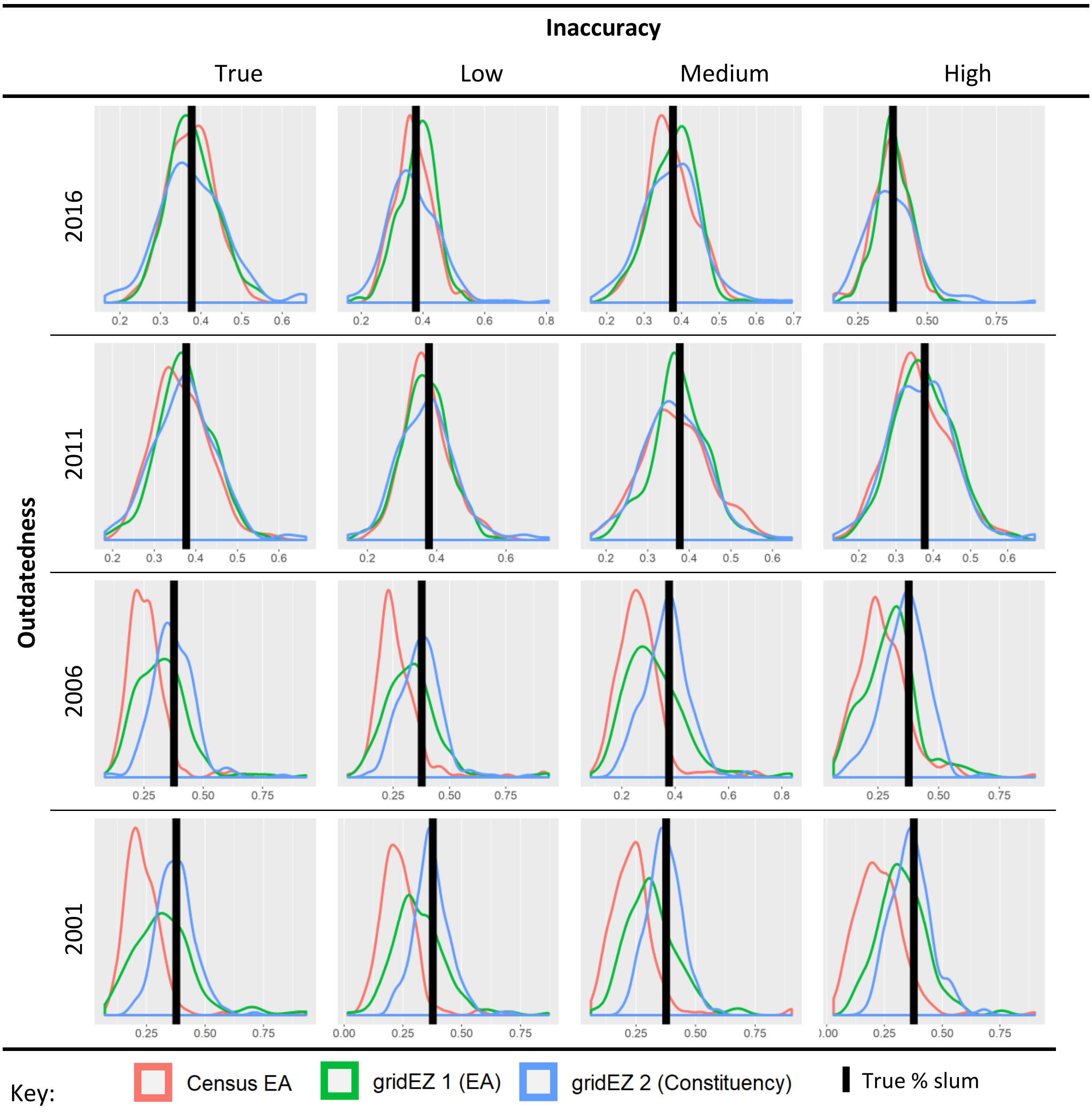


Figure 39. Percent of each sample comprised of slum households in the three sample frames, across all 16 scenarios of outdated-inaccurate census

Furthermore, according to Groves and Lyberg's (2010) description of the Total Survey Quality framework, lack of within-urban stratification on slum/non-slum areas is a "relevance" problem because without the means to disaggregate survey results by meaningful sub-groups, decision-makers do not gain the insights they need to address urban inequalities and risks.

5.4.2.1 Estimated indicator means

Box-plots of indicators for a selection of outdated-inaccurate sample frames are presented in Figure 40, and box-plots for all indicators and scenarios are presented in Appendix F. In census EA samples drawn from outdated or inaccurate data, indicators were underestimated in the overall urban population (first column of Figure 40). This was because slum household locations closely align with census EA boundaries, and slum households were more likely to be missing from the frame than non-slum households. Together, this reduced the probability of selection of majority-slum EAs under scenarios of high census outdatedness or inaccuracy.

This pattern, however, was mitigated in gridEZ sample frames, and resulted in household indicator estimates closer to the true prevalence (first column of Figure 40). While EA-level census inaccuracies were maintained in the underlying gridded population estimates, the gridEZ algorithm increased the probability that slum households were selected by grouping cells into PSUs across EA boundaries. In the most outdated and inaccurate scenario (2001_H), so much of the gridEZ-constituency input population was misallocated to areas comprised of (missing) slum households, that the outdated-inaccurate samples produced indicator estimates that were as accurate as samples drawn from more accurate sample frames. This was not necessarily a strength of the Random Forest model, but rather could be viewed as a fortunate coincidence.

Note that overcrowding (sufficient living and sleeping space inside the dwelling) was slightly more prevalent among non-slum households in the urban Khomas simulated data (scenario 1 in first column of Figure 40). This figure is likely representative of reality. In cities worldwide that face housing crises, overcrowding is an issue that effects households across economic classes (Bashir, 2002; Brown, 2003). Windhoek, unlike other fast-growing cities such as Dhaka, Bangladesh or Lagos, Nigeria, is not constrained by natural geographic boundaries. In satellite imagery, I observed that between 2001 and 2016 nearly all slum households were added to the urban periphery, and so they likely did not face the types of building constraints that are present in other fast-growing LMIC cities bound by the sea other natural barriers.

Estimates generated for slum and non-slum households separately after sampling (second and third columns of Figure 40) tended to be unbiased except in census EA and gridEZ-EA samples for indicators that differed substantially between slum and non-slum households (i.e., percent of households with an unimproved toilet or non-durable floor). The variability in indicator estimates across the 200 draws was not substantially different for non-slum households, even when the sample frame was outdated or inaccurate (second column of Figure 40). However, slum household indicator estimates varied substantially more across the 200 draws, and indicator estimates became more varied under scenarios of increased outdatedness or inaccuracy do to relatively smaller sample sizes of slum households (third column of Figure 40). In cases where the true indicator value differed substantially between slum and non-slum households, census samples – and to a lesser extent gridEZ-EA samples – tended to overestimate slum indicators. A possible reason for this is that the census or gridEZ-EA sample frames gave an inaccurately low probability of selection, but those census EAs or gridEZ-EA units that were selected tended to have higher concentrations of slum households, among which outcomes tended to be "worse" than in slum households near or interspersed among non-slum households.

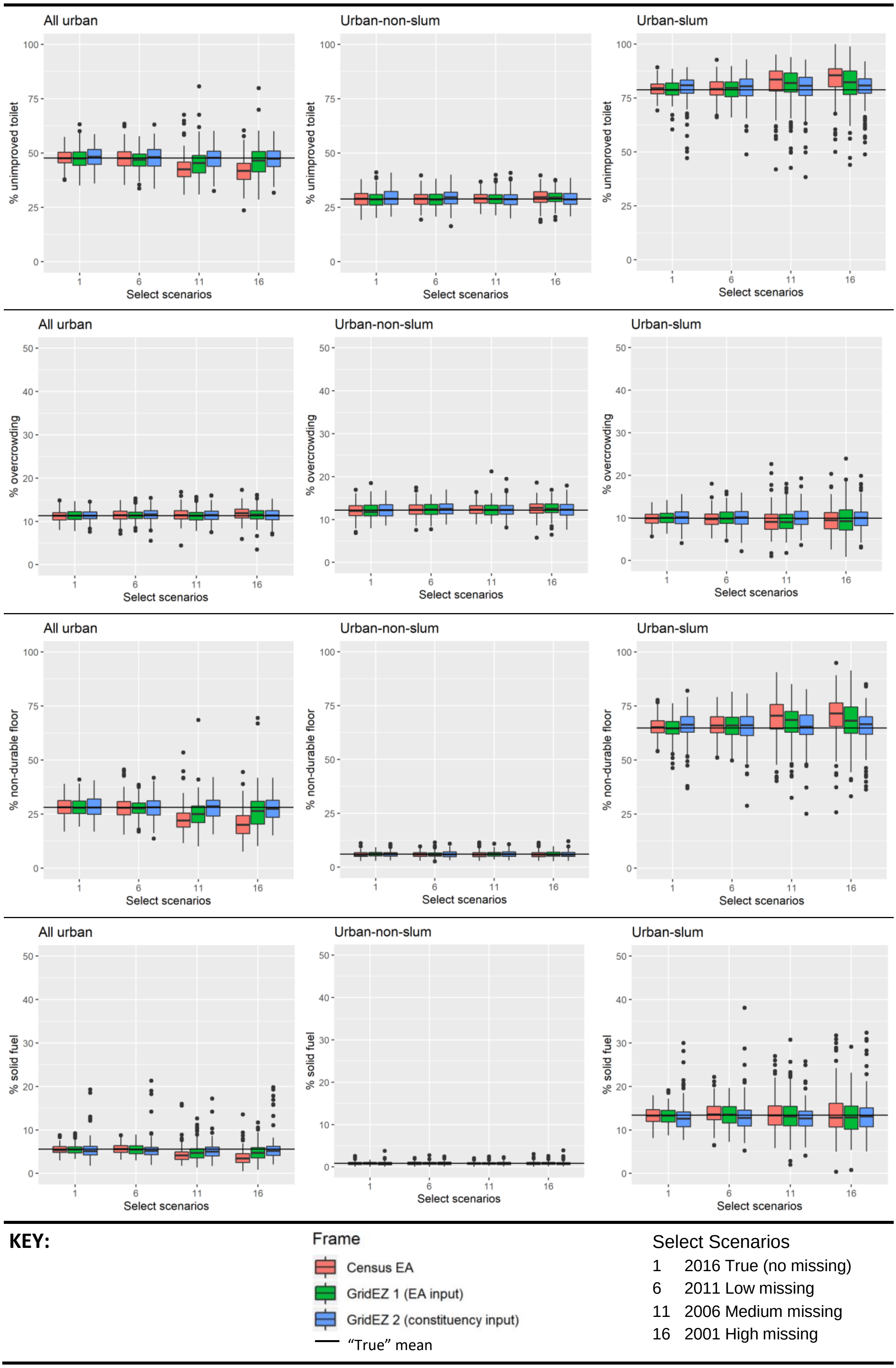


Figure 40. Box plots of household indicators estimated from 200 samples; by frame, outdated-inaccurate census scenario, and slum/non-slum households

5.4.2.2 Estimated indicator error

Table 31 summarises sampling error as measured with RMSE in select census scenarios, and average RMSE error across all census scenarios for the four household indicators. RMSE for the three sample frames and all census scenarios are presented in Appendix G.

When the best-case scenario 2016 "true" sample frames were used, samples drawn from gridEZ-EA units were roughly as accurate as samples drawn from census EAs directly, and samples drawn from gridEZ-constituency units were less accurate, as would be expected (Table 31). However, this pattern inverted as sample frames were increasingly outdated or inaccurate, especially for indicators with large disparities between slum and non-slum households. For example, under the least accurate 2001_H scenario (2001 census with high level of missing population), RMSE for percent of all urban household with an unimproved toilet were 8.3 in census EA samples, 7.3 in gridEZ-EA samples, and 5.3 in gridEZ-constituency samples, with a similar pattern when slum and non-slum households were disaggregated (Table 31). When these errors are averaged across all 16 scenarios, one or both of the gridEZ sample frames produced more accurate indicator estimates than the census frames (except overcrowding for which all frames averaged the same RMSE) (Table 31). Average RMSE for disaggregated slum/non-slum indicator estimates were mixed with census and/or gridEZ-EA samples tending to be more accurate when the sample frames were recent and more accurate, and gridEZ-EA and gridEZ-constituency tending to be more accurate when the sample frames were outdated or inaccurate (Table 31).

Table 31. RMSE in samples drawn from census EAs and two gridEZ sample frames, by indicator, select outdated-inaccurate census scenario, and slum/non-slum households

Select scenarios & 16 scenario average	**Census EA**			**gridEZ 1** (EA input)			**gridEZ 2** (Constituency input)		
	All urban	Urban non-slum	Urban slum	All urban	Urban non-slum	Urban slum	All urban	Urban non-slum	Urban slum
% unimproved toilet									
2016_True	3.9	3.8	3.4	4.4	3.6	4.2	4.9	3.8	6.4
2011_L	4.6	3.3	4.8	4.2	3.4	4.7	5.2	3.9	6.1
2006_M	7.2	3.1	8.8	6.7	3.3	8.4	5.3	3.6	7.3
2001_H	8.3	3.7	9.0	7.3	3.3	9.0	5.3	3.3	6.8
Average	**6.0**	**3.5**	**6.5**	**5.7**	**3.4**	**6.6**	**5.2**	**3.7**	**6.7**
% overcrowding									
2016_True	1.3	1.8	1.5	1.2	1.7	1.6	1.3	1.8	2
2011_L	1.3	1.6	1.8	1.3	1.6	2.1	1.4	1.7	2.2
2006_M	1.5	1.4	2.9	1.5	1.7	2.6	1.4	1.7	2.4
2001_H	1.6	1.6	3.0	1.6	1.6	3.6	1.5	1.8	2.7
Average	**1.4**	**1.6**	**2.3**	**1.4**	**1.7**	**2.5**	**1.4**	**1.8**	**2.3**
% non-durable floor									
2016_True	4.0	1.4	4.3	4.0	1.2	4.9	4.9	1.4	6.9
2011_L	4.6	1.2	5.8	4.0	1.3	6.0	4.9	1.5	7.1
2006_M	7.9	1.3	10.0	6.9	1.2	8.8	5.3	1.4	7.5
2001_H	10.1	1.5	11.1	8.3	1.3	10.4	5.3	1.4	7.7
Average	**6.7**	**1.4**	**7.8**	**5.8**	**1.3**	**7.5**	**5.1**	**1.4**	**7.3**
% solid fuel									
2016_True	1.1	0.4	1.9	1.1	0.4	1.9	2.1	0.4	3.3
2011_L	1.2	0.4	2.4	1.2	0.4	2.6	2.3	0.5	3.6
2006_M	2.2	0.4	3.5	2.0	0.3	4.0	1.8	0.4	3.1
2001_H	2.6	0.4	5.1	2.1	0.4	4.2	2.6	0.5	4.2
Average	**1.8**	**0.4**	**3.2**	**1.6**	**0.4**	**3.2**	**2.2**	**0.5**	**3.6**

5.4.2.3 Estimated indicator bias

As indicated in the boxplots in section 5.4.2.1, three of the four indicators (unimproved toilet, non-durable floors, and solid fuel) were underestimated in urban Khomas by outdated or inaccurate census EAs frames. In the 2001_H scenario, census sample frames underestimated the percent of households with an unimproved toilet by 5.9%, percent of households with a non-durable floor by 7.9%, and percent of household using solid fuel by 1.9% (Table 32). While gridEZ-EA and gridEZ-constituency also underestimated these indicators, the degree of underestimation was less severe: households with an unimproved toilet was underestimated by 1.3% by gridEZ-EA samples and 0.4% by gridEZ-constituency samples; households with non-durable floors was underestimated by 1.7% in gridEZ-EA samples and by 0.7% in gridEZ-constituency samples; and household using solid fuel was underestimated by 0.7% in gridEZ-EA samples and 0.1% in gridEZ-constituency samples (Table 32).For reasons discussed above (section 5.4.2.1), there was little bias in the overcrowding estimates across sample frames.

Table 32. Bias in samples drawn from census EAs and two gridEZ sample frames, by indicator, select outdated-inaccurate census scenario, and slum/non-slum households

Select scenarios & 16 scenario average	Census EA			gridEZ 1 (EA input)			gridEZ 2 (Constituency input)		
	All urban	Urban non-slum	Urban slum	All urban	Urban non-slum	Urban slum	All urban	Urban non-slum	Urban slum
% unimproved toilet									
2016_True	0.0	-0.1	0.4	0.0	0.0	-0.1	0.4	0.4	0.9
2011_L	-0.2	0.1	0.3	-0.7	-0.2	0.3	0.1	0.5	0.8
2006_M	-4.8	0.2	3.0	-2.3	0.2	2.2	-0.5	0.1	0.5
2001_H	-5.9	0.9	4.9	-1.3	0.7	2.1	-0.4	0.1	0.7
Average	**-2.7**	**0.3**	**2.2**	**-1.1**	**0.2**	**1.1**	**-0.1**	**0.3**	**0.7**
% overcrowding									
2016_True	0.0	-0.1	-0.1	0.0	-0.1	0.2	0.1	0.1	0.1
2011_L	0.1	0.1	-0.2	0.2	0.2	0.0	0.3	0.4	0.0
2006_M	0.2	0.1	-0.7	0.0	0.1	-0.6	0.1	0.1	0.0
2001_H	0.5	0.4	-0.6	0.2	0.3	-0.4	0.1	0.1	0.0
Average	**0.2**	**0.1**	**-0.4**	**0.1**	**0.1**	**-0.2**	**0.2**	**0.2**	**0.0**
% non-durable floor									
2016_True	0.0	-0.2	0.5	0.0	0.1	-0.2	0.3	0.1	1.0
2011_L	0.0	-0.1	1.2	-0.4	0.0	0.8	-0.3	0.0	0.5
2006_M	-5.6	0.0	4.4	-2.8	0.1	2.6	-0.5	0.1	0.5
2001_H	-7.9	-0.1	5.3	-1.7	0.1	3.0	-0.7	-0.1	0.5
Average	**-3.4**	**-0.1**	**2.9**	**-1.2**	**0.1**	**1.6**	**-0.3**	**0.0**	**0.6**
% solid fuel									
2016_True	-0.1	0.0	-0.1	-0.1	0.0	-0.1	-0.1	0.0	-0.5
2011_L	0.1	0.0	0.5	-0.1	0.0	0.1	-0.1	0.0	-0.4
2006_M	-1.3	0.0	0.3	-0.8	0.0	0.0	-0.4	0.0	-0.7
2001_H	-1.9	0.0	0.2	-0.7	0.0	-0.3	-0.1	0.0	0.0
Average	**-0.8**	**0.0**	**0.2**	**-0.4**	**0.0**	**-0.1**	**-0.2**	**0.0**	**-0.4**

5.4.3 Recommendations

In this analysis of a typical LMIC urban setting, gridded population sampling from a gridEZ-EA or gridEZ-constituency frame generally produced more accurate estimates for the overall urban population than a census EA sample frame when the sample frame was more than ten years old and/or inaccurate. Like most LMIC household surveys, the surveys evaluated here were not stratified within urban areas by deprived (e.g., slum) and not deprived (e.g., non-slum) areas. In this common design, gridded population sampling from gridEZ, or similar, units produce slightly more accurate indicator estimates within slum and non-slum sub-populations because each gridEZ PSU is likely to contain a greater diversity of household types than census EAs, which tend to follow homogenous neighbourhood boundaries. A key finding from this chapter is that urban household survey results based on a sample frame that is a decade or more old are biased toward "better" outcomes, especially when there are sharp disparities in outcomes between deprived and not deprived households, and when deprived households comprise a smaller portion of the overall urban population. This bias toward "better" outcomes in urban indicator estimates has major implications for LMIC household survey accuracy and decision-making.

In LMICs, urban socioeconomic disparities are stark and increasing (UN-Habitat, 2003, 2016). Worldwide, health and social indicators are strongly correlated with poverty/wealthy, and these correlations tend to be stronger in settings of greater socioeconomic disparity (Ezeh *et al.*, 2017). Despite the fast growing number of poor populations in LMIC cities, the proportion of the population that is poor versus non-poor is not well understood, and likely varies by city. The UN-Habitat "slum household" definition is widely used to classify households that lack basic assets from survey or census data (UN-Habitat PSUP, 2016); however, these data cannot be used to draw maps that would help to stratify a future survey. Furthermore, "slum households" as defined by UN-Habitat include most or all of some city populations, and do not necessarily correlate with slum/deprived areas as defined by local residents (Dana R Thomson, Kuffer, *et al.*, 2020).

While some city authorities and academic groups have produced city-level maps of slum areas, these maps are based on varying definitions, and they are rarely available across all urban areas in a country (Kuffer, Pfeffer and Sliuzas, 2016; Mahabir *et al.*, 2018; Thomson, Kuffer, *et al.*, 2020). Furthermore, slum maps become outdated quickly due to the particularly dynamic nature of slums. Given that there are few, if any, countries with harmonized, national, routine slum mapping initiatives, virtually no DHS, MICS, LSMS, and other routine national household survey stratifies samples within urban populations. The 2005-06 and 2015-16 India National Family Health Surveys were rare exceptions, and they only stratified slum/non-slum areas in eight cities (IIPS and Macro International, 2007; IIPS and ICF International, 2017). Given that the average DHS sample frame was seven years old (section 2.3), I suspect that a large portion of LMIC household surveys collected in recent decades have produced bias, likely overly cheery, estimates for urban indicators. If this suspicion is true, the implications are massive.

Biased household survey results which underestimate "bad" social and health outcomes mask the basic needs of the urban poorest, limit the ability of advocacy groups to raise awareness of the needs among the urban poorest, and undermine monitoring and evaluation efforts that target the urban poorest. The lack of maps identifying urban deprived areas is, in itself, an act of masking – even marginalizing - the needs of the urban poor. Until slums/deprived areas can be mapped routinely across all or most LMIC cities with reasonable accuracy, household survey planners will be unable to stratify surveys by this important sub-population, reinforcing the pattern of ignoring the urban poorest. The evidence presented in this chapter indicates that gridded population sampling could improve the representation of slum dwellers in household surveys in the absence of slum/deprived area maps.

I am hopeful that the increasing quantity and resolution of Earth Observation and Big data will enable routine, accurate mapping of slums and informal settlements across cities in the next five to ten years (Thomson, Kuffer, *et al.*, 2020; Kuffer *et al.*, 2020). When such maps become available, survey programme implementers will be faced with decisions about if, and how, to modify survey sample designs to produce separate estimates for rural, urban-slum, and urban-non-slum populations. Currently in many LMICs, urban areas are oversampled because either the urban population comprises a smaller portion of the national population, characteristics vary more within urban populations than within rural populations, or both. With an additional stratum (or strata) defined by slums, additional households may need to be sampled; though sample sizes within urban-slum and urban-non-slum areas might simultaneously decrease due to reduced variance within strata. The overall impact on sample size, cost, and effort would likely be moderate, but worthwhile to produce more accurate and disaggregated estimates that are appropriate for the essential monitoring, planning, evaluation, and research activities routinely performed with household survey data (Corsi *et al.*, 2012).

Chapter 6: Implementation

6.1 Overview

Survey implementation occurs after the sample frame has been developed (Chapter 4), the sample has been designed, and the primary sampling units (PSUs) have been selected (Chapter 5). Nearly all large-scale health and economic surveys in low- and middle-income countries (LMICs) use census enumeration areas (EAs) as the initial sample frame, and thus implementation necessarily requires two phases of fieldwork. The first phase of fieldwork is to list and map households (the mapping-listing phase), and the second phase is to interview sampled households, which usually occurs several months later (the interview phase).

Since the establishment of survey methods four decades ago, the mapping-listing phase has remained largely paper-based. This is in spite of the availability of satellite imagery, global positioning systems (GPS) and geographic information systems (GIS) for mapping, and availability of tablets and the internet for listing. Conversely, LMIC societies have been undergoing profound change related to globalization and rapid urbanization for at least two decades. The continued use of outdated survey tools and methods in contemporary LMIC contexts likely contributes to decay in survey accuracy, with vulnerable and mobile populations excluded in greater numbers than fixed, family households. Using the right tools for mapping and listing has multiple benefits including decreased costs, decreased field work time, and improved quality of the final sample frame due to improved experience for field staff and routine quality checks by supervisors.

This chapter builds on findings in previous chapters and evaluates the use of innovative gridded population surveys methods and tools in a complex, real-world context. In this chapter, I describe a feasibility analysis conducted under the Surveys for Urban Equity (SUE) study in Kathmandu, Nepal; Dhaka, Bangladesh; and Hanoi, Vietnam. The full SUE study protocol is published elsewhere (Elsey *et al.*, 2018). The focus in this chapter is on the largest of the three SUE surveys in Kathmandu. I describe the Kathmandu SUE study in section 6.2 including the methods and tools used to overcome threats to population exclusion. Section 6.3 describes the qualitative and quantitative methods I used to evaluate coverage and feasibility of gridded population sampling in Kathmandu. Finally, section 6.4 includes the results and discussion of the following three research questions.

6.1.1 Research questions

Research Question E: Is a gridded population sample feasible to implement in a complex, urban setting in a LMIC? Specifically, how does total cost, time, and staff skill mix in a gridded population survey compare a census-based survey?

Research Question F: What is the experience of mappers-listers implementing a gridded population survey using SUE field tools and methods?

Research Question G: What, if any, evidence is there that SUE tools and methods overcome unintentional population exclusion of vulnerable and mobile populations?

6.2 Surveys for Urban Equity (SUE) study

The SUE study was coordinated by University of Leeds (UK) in collaboration with HERD International (Nepal), ARK Foundation (Bangladesh), CIPRB (Bangladesh), and Hanoi University of Public Health (Vietnam) between 2016 and 2018. I coordinated one work package (activity) to implement and evaluate gridded population sampling field tools and methods. Not only did the SUE study evaluate tools and methods to improve the representation of the urban poor in household surveys, it also piloted questions on mental health and injuries for use in routine surveys in LMICs, and investigated techniques to improve the use of survey data by local authorities. The project was funded under a Research and Innovation grant by the British Medical Research Council with the following main activities:

i) identify and test questions on mental health and injuries
ii) test affordable and efficient novel methods to reduce bias in urban surveys
iii) explore alternative approaches to defining households, and measuring wealth
iv) develop data visualisation tools to support decision-makers' use of survey data

The SUE study piloted questionnaire modules about accidents and injuries, and depression and anxiety because these issues account for a large portion of the global burden of disease, particularly among the urban poorest, however these topics are rarely included in large-scale household surveys (Campbell and Campbell, 2007). Mental health and substance use are the fifth largest disorder in the global index of disability adjusted life years (DALYs), and the leading cause of years lived with disability, particularly among women (Whiteford *et al.*, 2013). Among mental health and substance use disorders, depression is the leading cause of disability globally, followed by anxiety (Whiteford *et al.*, 2013). Traffic accidents and other unintentional injuries such as falls, drowning, and fires are among the top ten burdens of disease globally, particularly among men (Haagsma *et al.*, 2016). Additionally, the SUE surveys collected and compared several measures of household wealth including the DHS assets-based wealth index and LSMS earnings and expenditures indicators to understand how these measures compare with SUE's qualitative findings in poor urban communities (Elsey *et al.*, 2018).

The study was approved by University of Southampton Ethics Review Committee (ref:26819), as well as University of Leeds Medical Ethical Review Committee (ref:MREC16-137), Bangladesh Medical Research Council (ref:BMRC/NREC/RP/2016-2019/317), Nepal Health Research Council (ref:1761), and Vietnam Medical Research and Ethics Committee (ref:324/2017/YTCC-HD3).

6.2.1 Roles in the SUE-Kathmandu study

I led activity (ii) to develop gridded population survey implementation methods and tools, and conduct a mixed methods study to evaluate the above research questions. I also provided input on activity (iii) to modify the household questionnaire member roster, allowing for classification and comparison of household membership according to DHS, MICS, and LSMS household definitions.

I provided survey implementation training in Kathmandu to HERD International survey team members Mr. Radheshyam Bhattarai, Mr. Rajeev Dhungel, and Mr. Subash Gajurel. Additionally, I observed and supported the HERD International survey team during their fieldwork. I also coordinated with qualitative researchers at HERD International, Ms. Sudeepa Khanal and Ms. Shraddha Manandhar, and the SUE Project Principal Investigator, Dr. Helen Elsey, to design a focus group with the mapping-listing staff after survey mapping and listing fieldwork was complete. This work was monitored and supported by HERD International Director, Dr. Sushil Baral.

I was not present for the survey implementations in Dhaka and Hanoi, and thus do not report results here. The Dhaka and Hanoi surveys were much smaller than in Kathmandu, and designed only to evaluate how the methods and tools designed in Kathmandu translated to other complex urban settings. Those two surveys were implemented by CIPRB and Hanoi University of Public Health, respectively, with training from the HERD International survey team following the Kathmandu survey.

6.2.2 Study setting

The cities selected for the SUE study represented a range of modern urban complexities and leveraged existing partnerships. All three cities face rapid, complex urbanization. Since 2010, the population increased by more than 4.0% in Kathmandu, 3.5% in Dhaka, and 5.2% in Hanoi (UN-

DESA, 2019). The pace of population growth in South Asia has particularly strained urban housing markets such that increasing numbers of people live in atypical arrangements such as hostels, or atypical locations including their place of work (UN-Habitat, 2016). Poorer households live in a variety of settings including physically isolated slum areas, as well as in dwellings located among brick-and-mortar buildings in economically heterogeneous neighbourhoods (UN-Habitat, 2016).

Kathmandu has a particularly complex and mobile population. Across the Himalaya mountain region, entire villages have traditionally migrated to lower altitudes during the winter, and both agricultural and non-agricultural seasonal migration to cities is common among residents of the Hill and Tarai (i.e., plains) regions (Gill, 2003). Kathmandu also experiences permanent/semi-permanent in-migration by rural and small-city residents for economic and educational opportunities, and to escape the detrimental effects of climate change on rural livelihoods (Tacoli, 2009).

The combination of rapid urbanization and unplanned infrastructure in Kathmandu made the 2015 earthquakes particularly devastating, and resulted in extensive short-term displacement and a rebuilding effort that continues today (Rimal et al., 2017). For all of these reasons, housing arrangements are particularly heterogeneous in Kathmandu (Figure 41). It is common for the owner of a building to occupy the top floor, rent the middle floor as one or two apartments, and rent the bottom floor to multiple lower-wage borders.

Credit: Hole in the Clouds *Credit: Wikimedia* *Credit: Wikimedia*

Figure 41. Examples of dense, complex housing arrangements in Kathmandu, Nepal

6.2.3 Innovative SUE methods and tools

I developed the following methods and tools utilizing innovative datasets with an aim to improve LMIC survey accuracy, with extensive feedback from Mr. Bhattarai, Mr. Dhungel, and Mr. Gajurel (Figure 42).

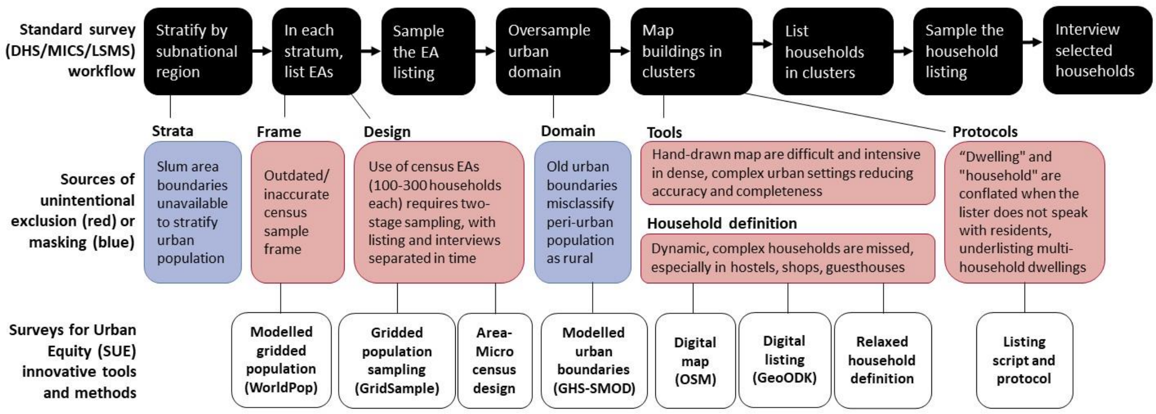


Figure 42. SUE methods and tools used to overcome unintentional exclusion and masking of vulnerable and mobile populations in household surveys
Adapted with permission from (Dana R. Thomson *et al.*, 2020)

WorldPop-RF sample frame (data). As described in Chapter 4, we used WorldPop-RF gridded population estimate in approximately 100 metre by 100 metre grid cells as our sample frame (WorldPop, 2019). At the time of this work in 2017, the WorldPop-Global dataset was not yet released. The WorldPop-RF estimates are derived with a machine learning approach that disaggregates population counts in larger areas (e.g., last census by district) to grid cells based on dozens of spatial covariates derived from publicly available satellite imagery and GIS data (Stevens *et al.*, 2015). As detailed in Chapter 5, the grid cells can be sampled directly, or combined into larger units, for household survey sampling.

GHS-SMOD urban boundary (data). We used the European Commission's Global Human Settlement GHS-SMOD layer to define the boundary of the Kathmandu metropolitan area. We considered use of municipality boundaries; however, "metropolitan" boundaries only covered the downtown area of Kathmandu, and adding "sub-metropolitan" boundaries would have resulted in a substantial sample of rural communities beyond the peri-urban reach. GHS-SMOD classifies 1 kilometre by 1 kilometre grid cells worldwide as "high density urban," "low density urban," "rural," or "unsettled/remote" based on a dataset of built-up areas derived from satellite imagery

(GHS-BUILT), and the GHS-POP gridded population layer (see section 5.3.1) (European Commission, 2017). Presented with metropolitan, sub-metropolitan, and GHS-SMOD boundaries, the HERD International team judged the GHS-SMOD dataset to best reflected the functional city boundaries of Kathmandu Valley.

GridSample (tool). At the time of this study, GridSample.org was not available, so we used the GridSample R algorithm to select samples from WorldPop-RF datasets. The free GridSample R package enables selection of primary sampling units (PSUs) with probability proportionate to size (PPS) from any gridded population estimate (see section 5.3.1) (Thomson *et al.*, 2016).

Area-Microcensus design (method). An area-microcensus sample was evaluated against the standard two-stage sampling design. In area-microcensus PSUs (15-20 total households), all households were invited to participate in an interview, while approximately one in every 10 households were sampled and invited to participate in two-stage PSUs (150-200 total households) (see 5.2.3). Area-microcensus designs have been used in a number of other gridded population surveys (Cajka *et al.*, 2018).

Relaxed household definition (method). The DHS, MICS, LSMS, and other large-scale surveys define a household as one or more people who are a usual resident, or who slept in the dwelling the previous night, and who share living arrangements and meals (see section 2.4.3) (Grosh and Munoz, 1996; ICF International, 2012a; UNICEF, 2013). These definitions are unclear about how, or whether, to include members of fluid "open" households defined by an in and out flow of individuals, often from rural areas, seasonal households and household members, and households that face housing insecurity and relocate often, all of which are common scenarios in South Asian cities (Oya, 2015). The SUE study relaxed the household definition to include all self-reported usual residents, as well as hostel dwellers and long-term occupants of guesthouses (defined as 7+ consecutive days and working, looking for work, or in the city for another purpose such as supporting someone in hospital), and street-sleepers who slept in the PSU the previous night. The questionnaire collected information about living arrangements, meals, and length of time in the dwelling to identify individuals who would not be counted in the household according to DHS, MICS, and LSMS household definitions.

OpenStreetMap map (data) and iD Editor (tool). We replaced hand-drawn paper field maps with digitally-generated paper field maps. To generate digital maps, we updated all buildings, roads, and pathways located in sampled PSUs in the crowd-sourced OpenStreetMap platform, using the iD Editor tool, which is integrated on the OpenStreetMap website (OpenStreetMap contributors, 2000). In the case that a building was intersected by a PSU boundary, we used the following rule: buildings intersecting west (vertical) and south (horizontal) boundaries were included, while buildings intersecting east and north boundaries were excluded, ensuring that each building and its residents could only ever belong to one PSU. We then exported all roads, buildings, and points of interest for Kathmandu Valley to a shapefile. Using ArcGIS 10.3, we created a map for each PSU showing the OpenStreetMap base layers, PSU boundary, and other PSU information such as centroid latitude-longitude coordinate. We used printed maps to record any changes to buildings, paths, and roads in the field. These paper maps were also very helpful to explain the survey to residents. Any modifications to the map were entered into OpenStreetMap using iD Editor in the office after fieldwork. Residential makeshift housing and tents were not entered in OpenStreetMap to prevent potential harm to residents (e.g., eviction, fines).

GeoODK listing (tool). We replaced the paper household listing form with a digital listing form in GeoODK (GeoMarvel, 2017), a free open-source data collection application that runs on a tablet. Several such applications are available including OpenMapKit (American Red Cross, 2017) and Survey Solutions (World Bank, 2020). We chose GeoODK because it allowed us to visualize satellite imagery and OpenStreetMap offline, and to record a manually placed latitude-longitude coordinate over each building. The same tool was used to administer interviews.

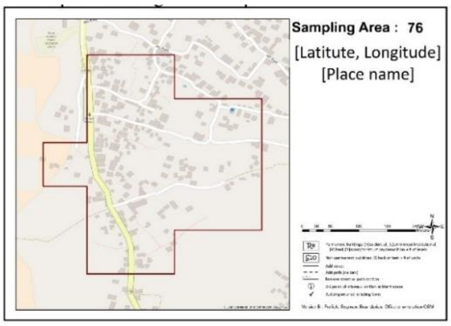
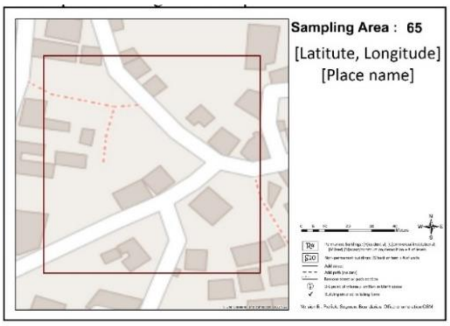
OSMAnd (tool) and MAPS.ME (tool). Several other applications were preloaded onto tablets to support field teams with navigation. MAPS.ME includes roads, buildings, and other features from OpenStreetMap for an entire country (or sub-region) and offers offline routing services (My.com, 2020). OSMAnd allows download of OpenStreetMap as well as satellite imagery to the tablet, plus the visualization of boundaries (PSUs in our case) and tablet location services (displaying the tablet location as a blue dot on the map) (OsmAnd BV, 2010). The offline satellite imagery required substantial storage space on tablets, so only imagery for PSUs were downloaded. MAPS.ME was used to navigate to PSUs, and OSMAnd was used to navigate within PSUs.

Household listing script (method). Clear protocols are not published by DHS, MICS, LSMS, or other largescale survey implementers to guide the household listing fieldwork (see section 2.4.3). To ensure that atypical households were not excluded (e.g., people who live at their place of work, multiple families who share an apartment), we trained listers to use a detailed script of questions at every building, and to seek information from neighbours in a standardized way when building residents were not home.

6.2.4 Study design

The HERD International team performed household listing and interviews in Kathmandu Valley between September and December 2017. The survey targeted 1200 households in 60 clusters to

Table 33. Summary of SUE Kathmandu survey design and methods used across two study arms, and household response rates by study arm

Design	**Two-stage**	**Area-Microcensus**
Coverage	Kathmandu Valley, general population	Kathmandu Valley, general population
Sample size	30 clusters, 600 households	30 clusters, 600 households
Cluster definition	Multiple, contiguous 100x100m cells with approximately 200 households	Single 100x100m cell with approximately 20 households
# clusters dropped & replaced	6	3
# clusters segmented	15	7
Mapping-listing	Mapping-listing team maps buildings, and lists dwellings and households	Mapping-listing team maps buildings, and lists dwellings. Interview team lists households on day of interview.
Mapping-listing staff	Undergraduate geospatial specialists, mostly male	Undergraduate geospatial specialists, mostly male
Interview	Interview team returns to sampled households in PSU approximately two months after household listing	Interview team returns to entire PSU approximately two months after dwellings (apartments) are listed. Interviewers perform household listing and interviews in all PSU dwellings.
Interview staff	Undergraduate public health specialists, even mix of male/female	Undergraduate public health specialists, even mix of male/female
Household response rate	96.8%	88.3%
Households met DHS & MICS def.	99%	90%
Example field map		

estimate key poverty, mental health, and injury outcomes (Elsey *et al.*, 2018). We randomized half of the clusters to an area-microcensus arm, and the other half to a two-stage arm, to compare the types of households and differences in outcomes in the different survey designs. In the area-microcensus arm, mapping-listing teams mapped buildings and listed dwellings (not households), while in the two-stage arm, the teams mapped buildings, and listed dwellings and households (Table 33). In the area-microcensus arm, the interview team *defacto* listed all households as they sought interviews with all households present in the PSU.

Although we targeted 60 PSUs, we sampled 20% (18) backup PSUs. A sample of 78 PSUs were drawn from the 2017 WorldPop-RF dataset with PPS using the GridSample R package (Thomson *et al.*, 2016). All 78 PSUs were initially designed to have approximately 200 households, or 820 people, using GridSample R's "growth" algorithm, which randomly adds neighbouring grid cells to an initial "seed" cell selected with PPS, achieving a target population or a maximum area parameter (Thomson *et al.*, 2016). I, along with the HERD International team, reviewed each PSU in ArcGIS 10.3 (ESRI, 2018) by overlaying it on satellite imagery, and we discarded PSUs with no inhabitable buildings (e.g., located over a factory or airport), and replaced it with a randomly selected backup PSU. Then we randomly assigned the 60 main PSUs to the area-microcensus or two-stage arm of the study. In the area-microcensus arm, the 100 metre by 100 metre "seed" cell became the new PSU boundary.

The HERD International survey team and I co-developed training materials (Thomson *et al.*, 2018). I delivered a mock training in English, colleagues provided detailed feedback, and I revised the slides and manuals in English. The HERD International team hired and trained 12 university student-employees specialising in geospatial methods to perform mapping-listing activities in September and October 2017. The mapping-listing training was delivered by the HERD International team in Nepali over one week and involved a mix of lectures, role-play, and practical activities. Mappers-listers were required to pass a field-based exam and desk-based test to qualify for the position.

As previously mentioned, the mapping-listing protocols used in area-microcensus and two-stage PSUs were slightly different. In both types of PSUs, mapper-listers mapped buildings; however, in area-microcensus PSUs they listed dwellings only, whereas in two-stage clusters, they listed all dwellings *and* households, and recorded the head of household name. The mapping-listing staff met at the office weekly, and on other days commuted from home directly to their assigned PSUs using provided stipends for transportation. At the end of each day in the field, mappers-listers called the HERD International survey team and submitted an electronic report. To minimize their commute times, we paired mapping-listing staff, in part, based on their home locations, and we assigned teams to nearby PSUs, though some PSUs remained far for any team.

In November and December 2017, the HERD International team hired 24 interviewers to visit sampled households and conduct a two- to three-hour interview. Interviewers were student-employees specialising in public health, and many had previous interviewing experience. The interviewers received a week of practice-based training which included map navigation and use of GeoODK for tablet data collection. The household questionnaire collected demographic, assets, savings, expenditures, migration, and injury information about everyone in the household or the household head. One adult in each household was randomly selected using the Kish table method (WHO, 2002) to complete an individual questionnaire which included mental health and migration questions. To be eligible, respondents were 18 years of age or older, a senior household member, and provided written informed consent.

6.2.5 SUE survey workflow

The focus of the rest of the chapter will be on the sample design, sample selection, and mapping-listing activities in the Kathmandu SUE survey in September and October 2017. This is because use of a gridded population sample frame and associated methods and tools effects sample selection and mapping-listing, but does not influence the interview tools and methods, nor the experience of the interview from the perspective of participants. An overview of the steps, roles, and tools are presented in Figure 43. At the time of this work, there were not any standards or guidance for implementing gridded population surveys, so I identified open-source tools from previous experiences and web searches, and pieced together a coherent workflow. In this workflow, the planning team was responsible for conducting an exercise to gauge the ideal grid cell size for the sample frame, sampling PSUs, assigning PSUs to one of the two study arms, and segmenting large PSUs (Figure 42). The mapping-listing staff were responsible for updating OpenStreetMap before fieldwork, conducting field-based mapping and listing activities, and updating OpenStreetMap periodically throughout fieldwork.

Sample frame scale exercise. To gauge accuracy of the 2017 WorldPop-RF sample frame and to ensure that “seed” cells were approximately the correct geographic size to capture 20 households in area-microcensus units, I developed the Sample Frame Scale Exercise (Appendix H). This exercise involved selection of ten 100 metre by 100 metre grid cells and ten 200 metre by 200 metre grid cells with PPS using the GridSample R algorithm. The HERD International team and I completed the exercise, and systematically assessing the WorldPop-RF estimates against high-resolution Google Earth satellite imagery (Google LLC, 2019) by counting what appeared to be single- and multi-family buildings. Through this exercise, we determined that the 2017 WorldPop-RF estimates appeared to be sufficiently accurate in nine out of ten 100 metre by 100 metre grid cells, and that most 100 metre by 100 metre cells likely had a minimum number of households to make area-microcensus sampling viable. We decided against the use of 200 metre by 200 metre cells in Kathmandu because this would have resulted in the need to extensively segment area-microcensus PSUs.

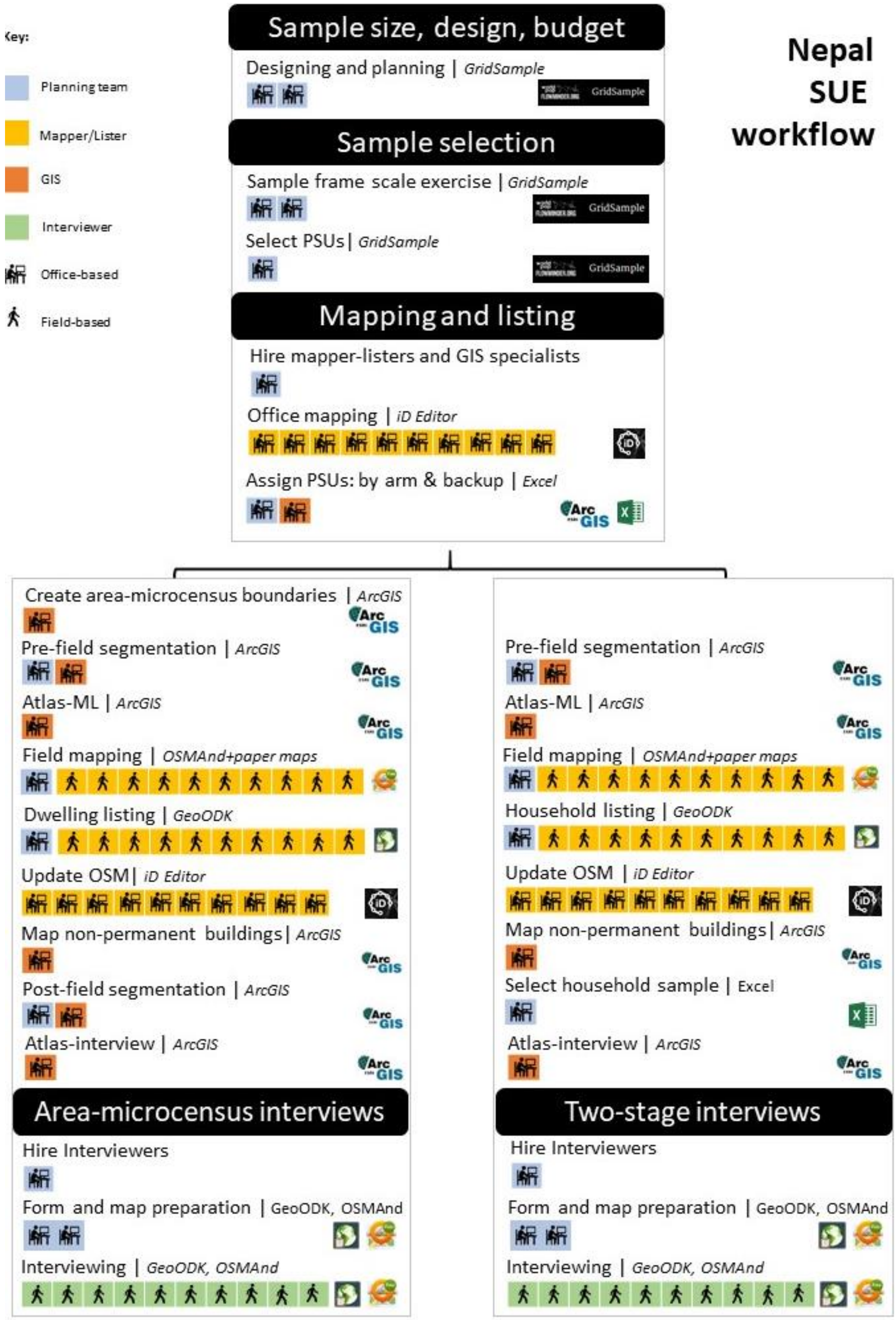


Figure 43. Overview of the workflow and tools used in both arms of the SUE Kathmandu survey

Adapted with permission from (Dana R. Thomson *et al.*, 2020)

Select PSUs. At the time of this study, the GridSample R package was the only gridded population sampling tool publically available. The only alternative was to develop an *ad-hoc* approach in a GIS (Dana R Thomson, Rhoda, *et al.*, 2020). The HERD International team members were not regular users of R or ArcGIS, so I drew the sample of PSUs, demonstrating and documenting my steps throughout the process. However, selection of gridded population PSUs can now be performed by non-GIS and non-R users with GridSample.org (see section 5.3.1).

Office mapping. We used OpenStreetMap (OpenStreetMap contributors, 2000) as the base layer in our field maps for several reasons. The first reason was to generate geographically accurate maps to promote accurate and timely navigation by the mapping-listing and interview teams in dense, complex communities. The use of standard hand-drawn maps would have been laborious and likely resulted in inaccurate maps due to dense, irregular roads and buildings. Second, the majority of buildings and roads in Kathmandu Valley had been mapped in OpenStreetMap, thus we were able to save time and human resources whilst maintaining anonymity of PSU locations. Use of this protocol in an area not well mapped in OpenStreetMap could have revealed PSU locations to an attentive OpenStreetMap user.

Assigning PSUs and pre-field segmentation. After updating OpenStreetMap but before generating field maps, the HERD International survey team and I reviewed each PSU over recent satellite imagery in ArcGIS 10.3 to decide whether to drop PSUs lacking habitable buildings. Dropped PSUs were replaced with a randomly selected backup PSU. After reviewing all PSUs, we randomly assigned 30 main PSUs to the area-microcensus arm and 30 to the two-stage arm, and backup PSUs were retained in the case a PSU was inaccessible or was found to lack residents during fieldwork. During our review, we also determined whether the PSU clearly had more than the target 20 (area-microcensus) or 200 (two-stage) households, and segmented those PSUs into two or more areas of approximately equal population size, then randomly selected one segment for inclusion in the survey (see Figure 44 for example segment maps). Segmentation was performed manually following roads and property boundaries in ArcGIS 10.3 (ESRI, 2018). The number of segments created in each PSU were tracked and used to calculate sample probability weights.

Field mapping and listing, and updating OSM. In the field, mappers and listers worked in pairs, taking turns updating the paper map and entering dwelling/household listing information into GeoODK on the tablet. They were trained to follow a script to approach residents, introduce themselves, explain the purpose of the survey, and ask detailed questions to understand and record residents’ living arrangements (script available at Thomson *et al.*, 2018). Upon request, listers distributed a written description of the SUE survey with the phone number and name of the survey coordinator. Weekly, the team spent half a day in the HERD International office updating buildings and roads in OpenStreetMap, debriefing with the survey coordinator, and connecting with other teams to swap stories, strategies, and build camaraderie.

Post-field segmentation. To ensure that area-microcensus PSUs had approximately 20 households each, a post-field segmentation step was conducted in area-microcensus PSUs only. In this step, the survey coordinator reviewed and manually segmented any PSUs with more than 25 dwellings, and tracked the number of segments per PSU so that sample probability weights could be appropriately adjusted later.

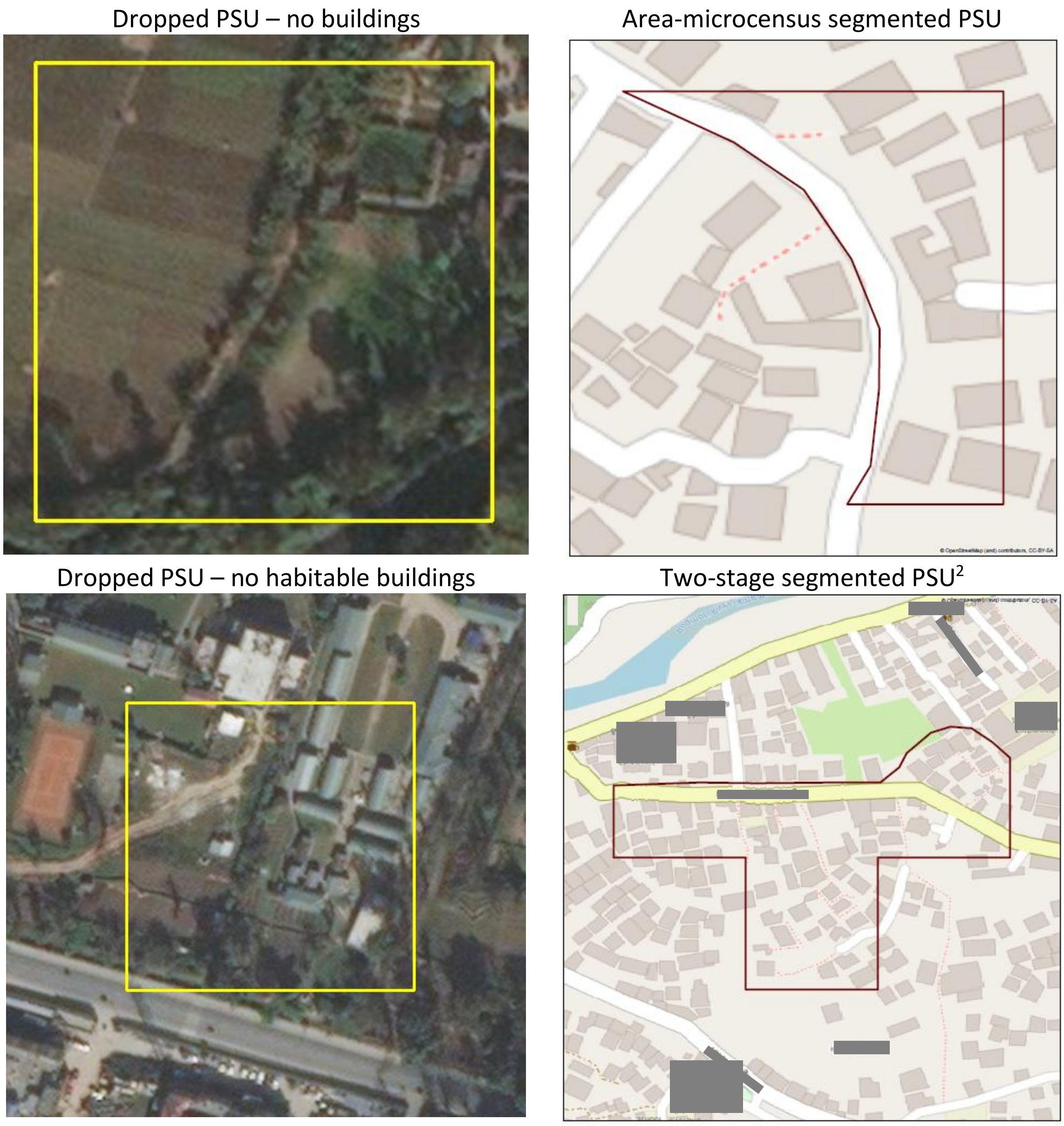


Figure 44. Examples of dropped and segmented primary sampling units (PSUs)

[2] Place name identifiers are masked to anonymize PSU location

6.3 Methods

I used a mixed methods approach to assess the feasibility of SUE gridded population survey methods and tools. This included participant observation with the HERD International survey team, and focus groups discussions and shadowing with hired mapping-listing staff. I also performed a simple quantitative analysis to compare types of households and outcomes in an area-microcensus versus two-stage survey design, and between the SUE household definition and the more constrained DHS/MICS household definition.

6.3.1 Qualitative

I was embedded with the HERD International survey planning team from July through September 2017, and conducted participant observation throughout that period, taking detailed notes at the end of most days. I also shadowed two of the six mapper-lister teams for a full day each, noting how they navigated with paper maps and tablet-based tools, their challenges, how they approached residents, how residents responded, and processes the team used to record data, particularly if their processes differed from protocols envisaged by me and the planning team. Occasionally, I provided feedback if it was requested, or asked questions to understand their thought processes.

A focus group discussion (FGD) was held with the mapping-listing team following their field work. With feedback from Dr. Elsey, Ms. Khanal, and Ms. Manandhar, I developed an FGD guide to elicit discussion about the office-based OpenStreetMap enumeration, field-based mapping and listing tools and methods, and overall workflow (see Appendix I). Ms. Khanal and Ms. Manandhar facilitated the FGD in Nepali over the course of two hours at the HERD International office. Ms. Khanal took hand-written notes and audio recorded the discussion, while Ms. Manandhar facilitated. Ms. Khanal transcribed the audio recording and translated the text to English, and Ms. Manandhar reviewed both the transcription and translation for corrections.

I performed a thematic Framework Analysis (Pope *et al.*, 2000; Gale *et al.*, 2013) on the English translated text in NVivo 11 (QSR International, 2018). This involved coding every line by 42 themes and summarizing results as a matrix – or framework – separated into three categories: positive/neutral experiences, challenges, and recommendations. The coding framework was developed iteratively with feedback from Dr. Elsey and qualitative researchers at University of Leeds and HERD International, and available in Appendix J.

6.3.2 Quantitative

To compare (a) household definitions and (b) survey designs, I calculated weighted percentages and means of key demographic and socioeconomic characteristics, and made comparisons with multinomial logistic regression at the 95% confidence level. This involved calculating sample weights according to the SUE household definition and the DHS/MICS household definition separately, following the formula presented in section 5.2.3, and normalising weights following the approach presented in section 5.4.1. The analysis of household definitions was constrained to the area-microcensus sample because, by definition, many of the atypical and vulnerable household members identified in the area-microcensus areas were *defacto* excluded in a two-stage design. The use of the SUE household definition in area-microcensus areas was assumed to represent a census of the "true" population, and thus was used as a reference population in statistical comparisons. The area-microcensus and two-stage arms were treated as strata in the analysis. I performed data management and analyses in Stata 14.0 (StataCorps LLC, 2020) using *svyset* commands to adjust for unequal probabilities of selection due to segmentation and stratification, and Taylor-linearized variance estimates for clustering of observations within PSUs.

Cost and time estimates were made for the area-microcensus survey and two-stage survey, separately. These estimates were made in collaboration with the HERD International finance team and survey coordinator. All expenses charged to the SUE project were considered, and the following assumptions were applied. I excluded the planning team salaries during the initial three weeks because we performed one-time activities including researching and piloting potential tools, fleshing out the basic workflow, and developing training materials that would not be repeated in future surveys. However, I included costs associated with normal survey planning tasks such as sample size calculation, preparing the sample frame, and selecting PSUs.

Salary estimates were based on number of person-days required to complete tasks in the area-microcensus and two-stage PSUs, respectively. Assuming that the planning team would perform the same tasks in either survey design, the only variable in costs between the two survey design arms was the total length of the survey period. Other costs related to ethics review, materials, and equipment were expected to be the same in an area-microcensus and two-stage survey.

6.4 Results and discussion

Multiple results are presented and discussed in this section including effect of household definitions and sample design on survey results; assessment of time, cost, and skill mix across sample designs; observations of the planning and field teams; and focus group discussions with the mappers-listers about their experiences using SUE tools and methods in the field.

6.4.1 Sample characteristics and design effects

In the area-microcensus sample, applying the DHS/MICS household definition resulted in exclusion of 10% of households (unweighted) compared to the SUE definition (Table 34). The treatment of survey arms as strata in the analysis meant that sample probability weights were larger in the two-stage arm because clusters comprised larger populations (SUE household definition mean weight: 1.673, range 0.298 – 5.524) than in the area-microcensus arm (SUE household definition mean weight: 0.347, range 0.157 – 0.985) (Table 34).

The root design effects (DEFTs) were larger in the area-microcensus PSUs than two-stage PSUs for demographic indicators using the SUE household definition (e.g. married: 2.13 vs. 1.23), but smaller for slum household (2.18 vs 2.40), migrant status (1.48 vs. 2.96), and education (2.95 vs 3.99) indicators (Table 34). The same pattern held using the DHS/MICS household definition (Table 34). Although one might expect larger design effects in the area-microcensus PSUs, because near neighbours are thought to be more similar than distant neighbours, the smaller DEFTs for slum, migration, and education indicators might indicate more complete coverage of heterogeneous neighbourhoods than typical two-stage samples. Smaller design effects with use of the SUE household definition compared to the DHS/MICS household definition might further indicate better identification of atypical and “hidden” households (Table 34).

6.4.2 DHS/MICS versus SUE household definition

Focusing on household definitions in the area-microcensus sample, nearly half of single adult households (46.9%, $p<0.001$) and sizable portions of migrant-headed households (6.7%, $p=0.016$), non-married (8.5%, $p=0.001$), unemployed (10.5%, $p=0.001$), and studying (14.3%, $p=0.003$) adults would have been excluded from the area-microcensus sample by using the DHS/MICS definition (Table 35). Results indicated that disabled adults might have been disproportionately excluded from the DHS/MICS household definition, though the sample size was quite small (9.3%, 0.009, weighted difference of 1 person) (Table 35).

Table 34. SUE Kathmandu sample characteristics and design effects

Adopted with permission from (Dana R. Thomson *et al.*, 2020)

	Two-stage		Area-microcensus	
Clusters				
Dropped and replaced	6		3	
Segmented	15		7	
Households				
Sampled - SUE	581		599	
Sampled - DHS/MICS (% of SUE definition)	578 (99%)		538 (90%)	
Household response rate	581/600 (96.8%)		599/678 (88.3%)	
Sample weights	***Mean (range)***		***Mean (range)***	
SUE	1.673 (0.298 - 5.524)		0.347 (0.157 - 0.985)	
DHS/MICS	1.581 (0.300 – 5.283)		0.346 (0.152-0.953)	
Design effects (SUE HH def.)	***Mean/prop. (SE)***	***DEFT***	***Mean/prop. (SE)***	***DEFT***
HH size	3.9 (0.111)	1.53	3.4 (0.137)	1.97
HHs per dwelling	1.0 (0.011)	2.11	1.9 (0.433)	4.20
HHs per PSU	19.5 (0.173)	4.42	24.9 (2.691)	5.40
Residential building	0.734 (0.023)	1.27	0.682 (0.075)	3.95
Nuclear family	0.517 (0.017)	0.83	0.439 (0.032)	1.56
Slum household	0.204 (0.040)	2.40	0.146 (0.031)	2.18
Migrant (head of HH)	0.700 (0.056)	2.96	0.780 (0.025)	1.48
Married	0.675 (0.014)	1.23	0.663 (0.026)	2.13
Employed full-time	0.459 (0.022)	1.82	0.486 (0.028)	2.21
Male 18+	0.371 (0.013)	1.34	0.416 (0.022)	2.02
Secondary+ education	0.495 (0.042)	3.99	0.528 (0.032)	2.95
Design effects (DHS/MICS HH def.)	***Mean/prop. (SE)***	***DEFT***	***Mean/prop. (SE)***	***DEFT***
HH size	3.8 (0.107)	1.47	3.4 (0.104)	1.48
HHs per dwelling	1.0 (0.011)	1.01	1.5 (0.088)	2.50
HHs per PSU	19.5 (0.174)	4.41	23.4 (1.533)	3.86
Residential building	0.734 (0.023)	1.27	0.718 (0.069)	3.55
Nuclear family	0.517 (0.018)	0.84	0.461 (0.026)	1.21
Slum household	0.215 (0.042)	2.44	0.170 (0.033)	2.02
Migrant (head of HH)	0.699 (0.057)	2.96	0.768 (0.024)	1.34
Married	0.675 (0.014)	1.23	0.678 (0.026)	2.09
Employed full-time	0.460 (0.022)	1.78	0.501 (0.028)	2.15
Male 18+	0.373 (0.014)	1.35	0.408 (0.023)	2.01
Secondary+ education	0.498 (0.042)	4.03	0.529 (0.032)	2.83

Table 35. Percent of households or individuals in the area-microcensus arm excluded by use of the DHS/MICS household definition versus SUE household definition

Adopted with permission from (Thomson, Bhattarai, *et al.*, 2020)

Indicator	N-wgt all	N-wgt DHS/MICS only	Wgt % excluded by DHS/ MICS	p-value†
Households				
Configuration				
Single adult	22	12	46.9	<0.001
One woman with children	10	10	0.0	<0.001
Nuclear family	91	91	0.6	Ref.
Other family *	73	73	0.6	0.906
Non-family	13	13	0.0	<0.001
*Slum household ** *				
No	171	163	5.1	Ref.
Yes	37	35	6.4	0.835
Migration status (head)				
Non-migrant	46	46	0.3	Ref.
Migrant	162	151	6.7	0.016
Adults 18+				
Marital status				
Not married	184	169	8.5	0.001
Married	364	355	2.3	Ref.
Employment status				
Full-time employed	267	262	1.6	Ref.
Part-time, underemployed	10	10	0.0	<0.001
Unemployed	27	24	10.5	0.001
Retired	20	19	1.9	0.839
Homemaker	123	122	1.5	0.860
Disabled “unable to work”	17	16	9.3	0.009
Student	82	70	14.3	0.003
Missing	2	0	100.0	<0.001
Individuals				
Gender and age group				
Male <12	55	54	1.4	0.139
Female <12	48	47	1.6	0.291
Male 12-17	31	30	4.9	0.822
Female 12-17	32	31	3.4	0.442
Male 18+	297	280	5.7	Ref.
Female 18+	251	244	2.8	0.203
Level of education				
Less than primary	171	163	4.7	0.733
Primary	124	118	4.6	0.711
Secondary+	377	362	3.9	Ref.
Missing	42	42	0.0	<0.001

* includes living with servants and/or extended family, sometimes with non-family household members
** defined as lacking improved water, improved sanitation, a durable structure, sufficient sleeping space (based on DHS/MICS household member definition), or insecure tenure
† multinomial logistic regression
N-wgt – weighted count

6.4.3 Area-microcensus versus area-microcensus design

Table 36 provides insight about the combined effect of exclusion due to survey design and household definition. The area-microcensus and two-stage arms of the study represented independent random samples that may have resulted in a mix of different households by chance. However, statistically significant differences in number of people per dwelling, building types, and household configuration followed expected patterns, which suggests that two-stage sample designs oversample residential family households and under-sample single adults residing in shared dwellings and mixed commercial-residential buildings.

Using the DHS/MICS household definition, the average number of residents per dwelling (living space) was 3.9 in two-stage PSUs compared to 5.0 in area-microcensus PSUs ($p<0.001$), suggesting that multi-household dwellings were easier to identify and record in area-microcensus units. Relatedly, non-family household configurations were less likely to be recorded in two-stage PSUs than area-microcensus PSUs (1.9% vs 6.3%, $p=0.029$). In two-stage PSUs, however, a larger portion of buildings were shacks/tents compared to area-microcensus PSUs (3.8% vs 0.7%, $p=0.009$) (Table 36). This is likely because the two-stage arm had much larger PSUs, and systematic sampling of households increased the chances of sampling informal settlements across the Kathmandu Valley. Despite these difference, no other statistically significant differences were detected among a select set of demographic and socioeconomic indicators.

However, when outcomes in area-microcensus PSUs using the relaxed SUE household definition were compared with typical two-stage PSUs using the DHS/MICS household definition, additional differences were detected. In the area-microcensus-SUE sample, the number of individuals per dwelling increased (5.3, $p=0.001$), and the number of individuals per household decreased (3.4, $p=0.013$), compared to the two-stage-DHS/MICS scenario. This is likely because the number of single-adult households included in the area-microcensus sample doubled to 10.4% by using a relaxed SUE household definition ($p=0.040$). The SUE household definition included individuals who stay in atypical accommodations, including hostels (3.8%), guesthouses (0.1%), and street sleepers (1.0%) who together represented nearly 5% of the area-microcensus-SUE sample. These additional individuals were not necessarily all single adult men, as one might expect. The distribution of males and females across age groups, education levels, employment status, marital status, migrant status (of head of household), and slum households remained statistically similar to the distribution observed in the two-stage-DHS/MICS sample (all comparisons, $p>0.05$). Though, differences in the weighted percentages between the area-microcensus-SUE and two-stage-DHS/MICS samples do suggest that many of the omitted individuals in standard household surveys are migrant adult men.

Table 36. Comparison of key demographic and socioeconomic characteristics by survey design (area-microcensus vs two-stage) and household definition (DHS/MICS vs SUE)

Adopted with permission from (Dana R. Thomson *et al.*, 2020)

Indicator	**Two-stage DHS/MICS (ref.)**		**Area-microcensus DHS/MICS**			**Area-microcensus SUE**		
	N-wgt	Mean or Percent	N-wgt	Mean or percent	p-value	N-wgt	Mean or percent	p-value
Survey Metrics								
HH size	928	3.9	191	3.5	0.014	208	3.4	0.013
Dwelling size	928	3.9	191	5.0	<0.001	208	5.3	0.001
HHs per PSU	928	19.5	191	23.4	0.016	208	24.9	0.051
Households								
Building type								
Residential	681	73.4 %	137	71.8 %	Ref.	142	68.2 %	Ref.
Mixed	206	22.2 %	50	26.4 %	0.595	52	25.0 %	0.594
Commercial	6	0.7 %	3	1.2 %	0.447	2	1.2 %	0.450
Shack or tent	35	3.8 %	1	0.7 %	0.009	1	0.6 %	0.009
Hostel	0	--	0	--	--	8	3.8 %	<0.001
Street sleeper	0	--	0	--	--	2	1.0 %	<0.001
Guesthouse	0	--	0	--	--	0	0.1 %	<0.001
Configuration								
Single adult	42	4.5 %	11	5.8 %	0.256	22	10.4 %	0.040
One woman with children	29	3.2 %	10	4.9 %	0.093	10	4.7 %	0.096
Nuclear family	480	51.7 %	88	46.1 %	Ref.	91	43.9 %	Ref.
Other family *	360	38.8 %	70	36.8 %	0.600	73	35.1 %	0.603
Non-family	17	1.9 %	12	6.3%	0.029	13	6.0%	0.030
Slum household **								
No	718	77.3 %	157	82.5 %	Ref.	171	82.3 %	Ref.
Yes	210	22.7 %	34	17.5 %	0.341	37	17.7 %	0.360
Migration status (head)								
Non-migrant	280	30.1 %	44	23.2 %	Ref.	46	22.1 %	Ref.
Migrant	648	69.9 %	147	76.8 %	0.244	162	78.0 %	0.173
Adults 18+								
Marital status								
Not married	861	32.5 %	163	32.2 %	0.924	185	33.7 %	0.107
Married	1,786	67.5 %	344	67.8 %	Ref.	363	66.3 %	Ref.
Employed full-time								
No	1,430	54.0 %	253	49.9 %	0.253	280	51.1 %	0.430
Yes	1,217	46.0 %	254	50.1 %	Ref.	267	48.7 %	Ref.
Missing	0	--	0	--	--	1	0.3 %	<0.001
Individuals								
Gender and age group								
Male <12	334	9.4 %	52	7.9 %	0.149	55	7.7 %	0.089
Female <12	232	6.5 %	46	6.7 %	0.875	48	6.7 %	0.710
Male 12-17	170	4.8 %	29	4.3 %	0.287	31	4.4 %	0.275
Female 12-17	181	5.1 %	30	4.5 %	0.330	32	4.5 %	0.275
Male 18+	1,329	37.3 %	271	40.8 %	Ref.	297	41.6 %	Ref.
Female 18+	1,318	37.0 %	236	35.6 %	0.202	251	35.2 %	0.118
Level of education								
Less than primary	957	26.9 %	157	23.8 %	0.412	171	23.9 %	0.440
Primary	599	16.8 %	115	17.3 %	0.880	124	17.4 %	0.906
Secondary+	1,774	49.8 %	351	52.9 %	Ref.	377	52.8 %	Ref.
Missing	234	6.6 %	41	6.1 %	0.601	42	5.9 %	0.494

* includes living with servants and/or extended family, sometimes with non-family household members
** defined as lacking improved water, improved sanitation, a durable structure, sufficient sleeping space (based on DHS/MICS household member definition), or insecure tenure
† multinomial logistic regression
N-wgt – weighted count

6.4.4 Time, cost, skill mix

Based on the Kathmandu budget, I estimated that the area-microcensus gridded population survey on its own, with a target of 600 households in 30 PSUs, cost US$26,769, or US$45 per household. The comparable gridded population survey using a two-stage design cost US$35,284, or US$59 per household in Kathmandu (Table 37). The main difference was that mapping and listing in each two-stage PSU took approximately 2.5 times than an area-microcensus PSU.

Table 37. Comparison of estimated time and budget to perform the area-microcensus arm of the survey versus two-stage arm of the survey, with 30 clusters and 600 households each

Budget Item	Two-stage design		Area-microcensus design	
	Time	Cost USD	Time	Cost USD
Planning & Administration				
Salaries	75 days	9,240	60 days	8,006
Mapping-Listing-GIS	35 days ×		12 days ×	
Salaries, per diem	6 mapper-listers	7,641	6 mapper-listers	3,056
Materials	1 GIS specialist	291	1 GIS specialist	218
Interviews & Data Management				
Salaries, per diem	19 days ×	5,723	15 days ×	4,518
Materials, including pilot	8 interviewers	2,106	8 interviewers	2,106
Ethics review		1,998		1,998
Equipment				
Laptops / hard drives	6 computers	1,193	6 computers	1,193
Tablets	4 tablets	1,212	4 tablets	1,212
Overhead (20% direct costs)		5,786		4,367
TOTAL		35,284		26,769
Cost per household		59		45

Thirty PSUs is a small sample size for a household survey; national household surveys typically have more than 400 PSUs. Training and supervision costs would benefit from economy of scale in a larger survey, while other costs, such as transportation and accommodation, would increase due to distant PSUs. This means that area-microcensus sampling could result in a large cost savings compared to the standard two-stage design. However, there is a potential trade-off in statistical power.

Area-microcensus sampling can result in a larger design effect if near neighbours are more similar than distant neighbours, though heterogeneity in urban residents likely varies by city and over time (Elsey *et al.*, 2016). A simulation study of design effects in rural Namibia suggested that an area-microcensus survey with PSUs of 25 households would need to double the number of PSUs to achieve the same statistical power as a two-stage survey in the same population (Kools, 2018). An increase in sample size would increase costs. However, in a more heterogeneous setting like

Kathmandu, the number of additional area-microcensus clusters to achieve the same statistical power as a typical two-stage sample might be far less. Furthermore, if the number of households per PSU were reduced, it is possible that area-microcensus sampling with a gridded population sample frame and additional PSUs could cost the same or less than a two-stage survey with a census-based frame. The benefit of the area-microcensus design is potentially more accurate representation of vulnerable and mobile households, particularly single-adult households.

The skill mix required to plan and implement a gridded population survey using SUE innovative methods and tools was not much different from a standard household survey. The main difference was the skillset of the mapping-listing team. In a standard survey, mapping-listing staff are required, at a minimum, to be map literate and have a secondary school education (ICF International, 2012a). It is increasingly common in household surveys that mapping-listing team leaders are able to use a GPS unit and have basic GIS skills to interact with maps and printed satellite imagery of PSUs. To use SUE methods and tools, the mapping-listing staff should be comfortable using mobile technologies for data collection or navigation. For larger surveys, it might be helpful to have a dedicated GIS staff member to produce maps.

The skills sets of other staff in the SUE gridded population survey, including survey planners, trainers, and interviewers were essentially identical to a standard household survey, assuming that survey planners use GridSample.org for sample selection. If the GridSample R package is used, intermediate programming and GIS skills are required. Most standard household survey questionnaires are now administered on a tablet, so most planning teams have protocols to set up a server, set up tablets, and preload forms and reference data. We used similar protocols to prepare tables for a digital household listing, and install apps to aid navigation. The roles and skills of the interviewing teams are identical in both survey approaches.

6.4.5 Participant observation and shadowing

The HERD International survey team was well versed in finding innovative solutions to massive survey challenges. Several colleagues, for example, were responsible for conducting a complete building census and needs assessment in 26 districts in and around Kathmandu following the 2015 earthquakes, and the team routinely collects surveys in some of the remotest villages on earth in the mountains of the Himalayas. Dr. Sushil Baral ensured that team members had time away from other projects to work on this survey, and Mr. Bhattarai, Mr. Dhungel, and Mr. Gajurel regularly stayed late in the office with me to hash out workflows, test applications, and prepare for the SUE survey. I recorded over 50 photographs of our white-boarding sessions, the content of which is reflected in the SUE Planning Team and Mapper-Lister-GIS manuals (Thomson *et al.*, 2018).

Although I coordinated analysis of the SUE survey, the innovations developed in this study were very much a group effort, as visualised in a few photos from our work together (Figure 45).

Figure 45. Images from participant observation with the Kathmandu SUE survey planning team

Here, I highlight just one of many personally impactful observations. During the week of mapper-lister training, members of the planning team were assessing participants in ways that I had not considered. When we finished calculating exam results on the last day of training, Mr. Bhattarai suggested that we finalize mapper-lister hiring and team assignments over coffee at a local café. I had assumed that this would be a mechanical and quick process; not hiring candidates who performed poorly in assessments, and pairing higher-performing staff members with anybody who passed the assessments but needed more confidence or skills.

Instead, the planning team members brought printed ID cards (that would later be given to each hired mapper-lister) to the cafe, and used them to visualise and discuss potential team pairings (top-right image Figure 45). They started by laying out the ID cards of five very high performing candidates – these five people would be on separate teams. Then they spent more than an hour discussing interpersonal dynamics that they had observed throughout the training week, the cultural and language backgrounds of candidates, and even how residents in certain sampled neighbourhoods might perceive the ethnic identity of mapper-lister names. The aim of this exercise was to maximise team skill *and* enjoyment in the field. Considerable thought was put into pairing candidates who were most likely to provide each other with the support and focus that they would need to work through long, tiring days in the field. The planning team also aimed to put staff in situations where they were most likely to succeed by trying to minimise any predictable negative interactions (e.g., ethnic discrimination). The key insight for me was that high-quality, accurate survey data starts with happy, well-supported fieldworkers. Frustrated, emotionally exhausted survey field staff can undermine the cleverest survey design, tools, or methods.

Shadowing two of the six mapper-lister teams was also very insightful. It provided me an opportunity to observe the amount of time it can take to get one's bearings in the field, particularly in a setting as complex and dynamic as Kathmandu. I observed fieldwork in both an area-microcensus and a two-stage PSU, and the experiences were quite different. Not only was the two-stage PSU much larger, the interactions with residents took much longer. Gathering head of household name was particularly time-intensive in the two-stage PSU. Multiple times, the mapping-listing team was asked to wait for the landlord to be called on the phone, or to come downstairs from an apartment, to provide names of building residents because no one else in the building was willing to do so without the landlord's permission. This sometimes took 20 to 30 minutes, during which time one team member went ahead and continued mapping and listing households.

By visiting PSUs across diverse neighbourhoods and speaking with residents, mappers-listers discovered many informal businesses and lifestyles. For example, we came across a glass-blowing factory behind someone's home that manufactured smoking pipes and bongs. The factory staff let me step inside and snap a photo (Figure 46). Next to the factory floor/store room, I observed about a dozen bunkbeds in two small rooms where workers stayed semi-permanently. This was just one informal residence listed that day.

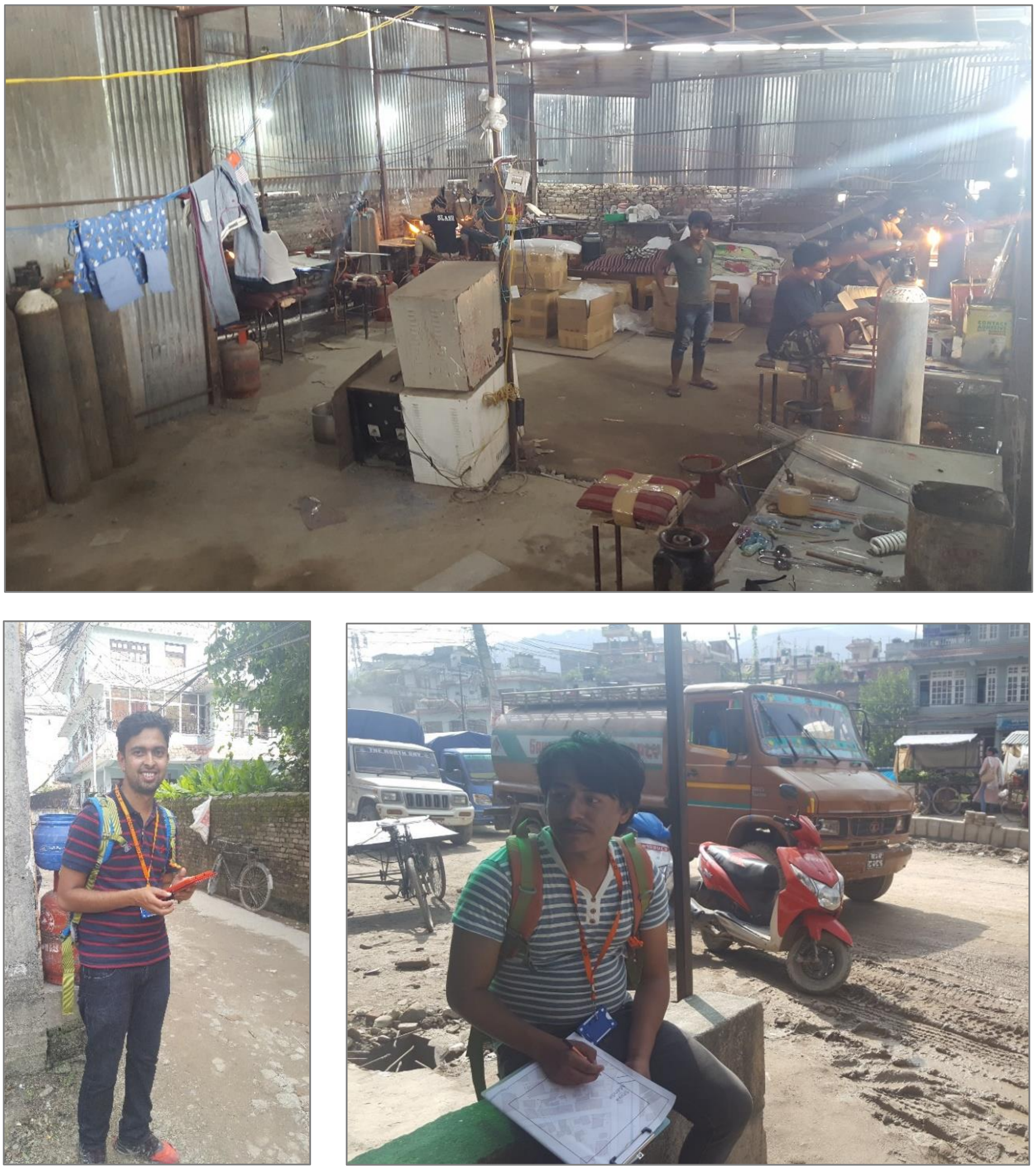


Figure 46. Images from shadowing mapper-lister teams during the SUE survey fieldwork

Shadowing field teams also allowed me to witness the routine rejections that fieldworkers face, and learn how they reacted and stayed motivated through these experiences. A large portion of residents were away from home at work during data collection, or busy with domestic work. Rejections by residents were sometimes dismissive, while other residents expressed outright anger at the mapping-listing team. The first team that I shadowed was yelled at three times by residents on the same day. After one such interaction, a man shouted down the road at us after we respectfully walked away, and I asked what the man was shouting. The team I was with said that he was complaining that household surveys are a waste of time and resources and never result in improvements for the neighbourhood or the city.

6.4.6 Focus group discussions (FGD) with mapping-listing staff

Feedback from the mapper-lister FGD was generally neutral or positive. The fieldwork, however, was not without limitations. The mapping-listing staff described their greatest challenge to be approaching residents. As geospatial specialists, they were generally adept at using technology, but had limited-to-no experience approaching members of the public on the street. The mappers-listers noted that role-play and practical activities helped prepare them for fieldwork; however, additional training about the overall survey aims would have been helpful to more clearly explain the survey's purpose to residents.

In FGDs, mappers-listers described working in pairs as essential because it provided them with "mutual support," allowing them to adapt to the moods and reactions of residents, interact in multiple languages, and to work faster and more accurately by supporting each other with navigation and recording details. Overwhelmingly, mappers-listers recommended that future gridded population survey teams be comprised of one geospatial expert who focuses on navigation and mapping, and one public health expert who is experienced at approaching residents to solicit household information.

Several mapping-listing staff felt that residents routinely omitted mention of neighbours who did not have official mortgages or rental contracts, presumably for fear of evictions or fines. Additionally, in wealthy "VIP" neighbourhoods, the mapping-listing staff said that they encountered substantial scepticism and non-response. Mappers-listers commuted from home to PSUs via bus, rickshaw, motorbike, and on foot. Most staff never travelled more than one hour to a PSU, however a team assigned to a peri-urban PSU spent three hours commuting each way due to the absence of buses or taxis. Several staff recommended hired vehicles to save time, especially in two-stage PSUs which could be quite large in peri-urban, less dense areas.

Mapping and listing staff reported different experiences in area-microcensus and two-stage PSUs. The two-stage PSUs were, by definition, at least ten times the size of area-microcensus PSUs; however, due to variability in the WorldPop-RF estimates, several two-stage PSUs were even larger resulting in extra days of work and more physical barriers to navigate, such as hills or rivers. In addition to being much larger, the two-stage PSUs required more information to be collected from residents than area-microcensus PSUs resulting in longer interactions and higher levels of scepticism among residents. Similar to my field observations, mappers-listers said that residents were generally happy to report number of dwellings per building; however, they were reluctant to specify the number of households per dwelling and to give household head names. In many two-stage PSUs, teams approached a person working in a shop on the ground level who gave number of dwellings in the building, but refused to give household-level information, and instead directed the mapping-listing staff to the building owner or landlord.

Mapping-listing staff faced some challenges with the tablet applications. While some challenges could have been averted with more training, other challenges were inherent to the tools and methods used. First, OpenStreetMap roads and buildings data were updated to the various tablet applications on different schedules resulting in different versions of the same map across applications. Mappers-listers updated roads and buildings in OpenStreetMap at the HERD International office before starting fieldwork, but the changes took several hours to appear in ArcGIS (from which paper field maps were printed), several days to appear in GeoODK (where building GPS points were collected as part of the listing), and up to 30 days to appear in OSMAnd and MAPS.ME (both used for navigation).

A second problem was the number of applications that the mapping-listing staff were expected to use. It was time-consuming and confusing to switch between MAPS.ME to navigate to a PSU, OSMAnd to navigate within a PSU, and finally GeoODK to manually record a GPS point over each building. Additionally, the location precision within OSMAnd and GeoODK were poor, often showing a massive blue circle in which the tablet could be located, undermining the purpose of this location feature. Despite these challenges, mapping-listing staff strongly preferred these tools and methods to the described alternative of hand-drawn maps and paper listing forms.

6.4.7 Recommendations

By comparing an area-microcensus sample design with relaxed SUE household definition and a standard two-stage sample design with DHS/MICS household definition in Kathmandu, Nepal, I found evidence that standard household surveys unintentionally omit vulnerable and mobile households, which might affect the accuracy of survey estimates. Standard household definitions, and the process of two-stage sampling, tend to result in listings of settled, family households, whereas area-microcensus-SUE surveys were able to identify more migrant single adults, hostel dwellers, people staying long-term in guesthouses, and street sleepers. However, without slum area boundaries to stratify urban area, SUE survey methods were unable to adequately measure shack and tent dwellers.

A key finding from this study is that the person performing the household listing, and under which circumstances, matters in modern urban settings. Although the same household definition, protocols, and tools were used by mapper-listers and interviewers who listed households in area-microcensus and two-stage PSUs, respectively, different types of households were listed in the two types of PSUs. While this difference could have occurred by chance, the focus group discussion and my field observations point to an alternative explanation. When interviewers performed the household listing in area-microcensus clusters, they had substantially more time to interact with residents than the mapping-listing teams in two-stage clusters, resulting in a more thorough listing. Furthermore, the interviewers in this survey had more experience and skills interacting with the public which may have influenced the quality of information provided by residents. In a two- to three-hour interaction, interviewers were able to build the trust necessary for residents to report atypical and “hidden” neighbouring households in area-microcensus PSUs; while these types of vulnerable households were likely not reported to mappers-listers in brief five to 15 minute interactions in two-stage PSUs.

The mappers-listers believed there was under-reporting of “hidden” lower-income households in shared or inconspicuous dwellings, especially in two-stage PSUs. Without any other way of deducing the existence of these households, they were omitted from the sample frame. While mappers-listers also reported that many “VIP” households refused to speak with them during the mapping-listing activity, “VIP” households were still partially listed and included in the sample frame because their dwellings were visible, and neighbours or guards could often provide missing information (i.e., number of households in the dwelling, and head of household name). A systematic exclusion of “hidden” households from the sample frame would have resulted in estimates biased toward middle-class and wealthier households due to coverage error among vulnerable or mobile households. One can also presume that both vulnerable/mobile and “VIP” households were less likely to respond to the survey than middle-class residents, in which case the sample would under-represent socio-economic extremes in the population. While sample weights (which include non-response rates) would duly increase variance estimates around indicators, an absence of the lowest and highest indicator values in the sample could still result in inaccurately narrow variance estimates around indicators. However, given the demographics and context in Kathmandu with far more vulnerable and mobile households than wealthy households, these errors were likely skewed toward wealthier households.

As LMIC urban contexts are increasingly defined by complex living arrangements, mobile, and temporary residences, there might be a need to move the household listing responsibility to interviewers using an area-microcensus survey design. Others argue that our current household definitions developed for censuses and surveys in the 1980s when population were majority rural are no longer suitable in complex LMIC urban settings; rather, individuals and communities (small areas) are more appropriate units of measurement (Campbell, 2017).

Another key finding is that tent and shack dwellers comprised a small percentage of households in both arms of the study, particularly in the area-microcensus arm. Households living in tents and shacks were more prevalent in two-stage PSUs because two-stage PSUs were substantially larger than area-microcensus PSUs, and thus had a greater probability of covering a diversity of neighbourhood and building types. However, the only way to ensure representative survey samples of shack/tent dwellers and other vulnerable populations concentrated in slums – in both area-microcensus and two-stage designs – is to treat slum/non-slum urban areas as strata (Lilford *et al.*, 2017).

The Kathmandu SUE survey had some technical and practical limitations. At the time the sample was designed and selected, there were no ready tools or algorithms to group grid cells into sampling units of similar population size. The gridEZ algorithm was later developed, which now performs this task to support household surveys and other similar fieldwork (Dooley, 2019). Instead, we used the published GridSample R package which “grows” PSUs after “seed” cells are selected. This means that sample weights in area-microcensus PSUs were consistent with standard household surveys because the “seed” cell was used as the PSU boundary. However, in two-stage clusters, the population of the “grown” PSU used in sample weights, was different from the actual probability of selection. To more accurately calculate weights for the two-stage clusters, an adaptive sampling weight could theoretically be applied; however, with very large areal constraints on PSU “growth,” the number of permutations, and thus number of terms required in an adaptive sample weight calculation approaches infinity and is not feasible to calculate.

A practical limitation of the field work was that response rates in area-microcensus PSUs were lower than in two-stage PSUs. This may have been due to the greater proportion of vulnerable and mobile households measured in area-microcensus PSUs; and those types of households may have been less willing to participate or were more likely to be absent during the daytime. This finding is contrary to a common concern about sampling methods which place interviewers in charge of household sample selection. Survey practitioners generally expect higher response rates to be recorded when interviewers decide which household to include in the survey because interviewers are assumed to omit difficult or undesirable households (UNSD, 2005b). I do not suspect this was a problem in the Kathmandu SUE survey for three reasons. First, a dwelling listing had been performed by a mapping-listing team before interviews which ensured that at least one household response was recorded for each listed dwelling. Second, the same interviewers worked in both area-microcensus and two-stage PSUs, likely resulting in consistency in practices across the PSU types. And third, the training, supervision, and support provided by the HERD International staff was outstanding, and likely motivated interviewers to adhere to the protocols.

These findings are from just one city, and may have differed in another city, or at a national scale. Future information about time, cost, skill mix, and experiences of staff would be needed to generalize about gridded population surveys, or to fairly compare area-microcensus and two-stage designs. This study, however, does suggest that gridded population sampling is a viable alternative to census-based sampling using either a micro-census or two-stage sample design. The time, cost, and skill mix requirements for both survey arms in Kathmandu were comparable, or cheaper, than other standard surveys implemented by the HERD International team on other projects.

Chapter 7: Conclusions

7.1 Key findings and implications

This thesis evaluates the accuracy and feasibility of gridded population sampling, focusing on the ability of gridded population sampling to represent the poorest and most vulnerable members of society in household survey data. In Chapter 2, I detailed the dramatic rates of urbanisation that have transformed low- and middle-income countries (LMICs), particularly over the last two decades, and how the continued use of survey methods and tools – developed for rural populations four decades ago – are likely masking or omitting the urban poor from household survey data. I hypothesised that poor, vulnerable, and mobile populations were disproportionately omitted from the census sample frames, and that existing survey methods favour long-term, stably-housed, family households. In Chapters 4, 5, and 6, I tested these hypotheses.

My research questions, however, required a highly detailed and accurate population dataset with information about household locations and outcomes in a typical LMIC urban setting. Unable to find such a dataset, I simulated a realistic population in Chapter 3 for Khomas, Namibia, a region that is 95% urban. I also systematically reviewed the literature to identify rates of under-counts among rural, urban slum, and urban non-slum households in LMIC censuses. I then used this information to simulate 15 realistic scenarios of outdated, inaccurate census populations from the "true" simulated population.

In Chapter 4, I used the 16 simulated datasets to create and evaluate cell-level accuracy in a gridded population dataset. After reviewing existing approaches to modelling gridded populations, I chose to replicate one of the most accurate and best documented approaches by the WorldPop team. I found that outdated, inaccurate census data had little effect on cell-level accuracy in WorldPop gridded population datasets. Instead, the main source of WorldPop cell-level inaccuracy was use of average population density at a spatially aggregated scale to make estimates of population density at a much finer geographic scale. This error resulted in gross under-estimates in the most densely populated grid cells (where slum dwellers were more likely located) and over-estimates in less dense cells. Although the grid cell population totals were underestimated across urban Khomas, I found that the distribution (percent) of the population across cells was reasonably accurate, within +/- 20% the "true" population distribution.

In Chapter 5, I selected 200 household samples from each of the simulated censuses and gridded population datasets derived from these simulated censuses. I limited the study area to urban

Khomas because the population totals and distributions measured in urban versus rural Khomas in Chapter 4 exhibited very different results, and because 95% of the "true" simulated population was located in urban Khomas. I compared repeated samples from (1) census sample frames, defined by actual census EA boundaries with minor modifications to account for urban expansion between 2011 and 2016, (2) gridded population data derived from census EA-level inputs, and (3) gridded population data derived from constituency (second-level administrative) inputs. To prepare the gridded population data for sampling, I used the publicly available gridEZ algorithm to group cells into units with a target population and a maximum area (similar to census EAs). I compared estimates of four household outcomes across the three sample frames and 16 census scenarios with the "true" simulated population, and found that gridded population samples derived from either census EA- or constituency-level data resulted in more accurate indicator estimates when the sample frame was more than ten years old or inaccurate.

A key feature of the simulated data in Khomas was that non-slum (non-poor) households comprised a majority of the population (about two-thirds), and had "better" outcomes than slum (poor) households. This was by design, following the distribution of actual household characteristics measured in Khomas in the 2011 census. This population configuration led to a key finding about household survey inaccuracy in the simulated urban Khomas population, and likely in many other similar LMIC cities: in the presence of moderately outdated or inaccurate census data, census-based samples produce overly cheery statistics because poor households were missing from the sample frame, had "worse" outcomes, and comprised a smaller portion of the overall urban population.

The solution to this problem is stratification of urban surveys by deprived/non-deprived areas; however, these types of maps are largely absent, or inconsistent if they exist, in LMIC cities. In Chapter 5, I performed sub-analyses in slum and non-slum households after sampling and found that this disaggregation did not improve accuracy of slum estimates from census samples when the frame was ten or more years old or inaccurate. However, in all scenarios, including outdated or inaccurate underlying census, gridded population samples from gridEZ units resulted in more accurate overall urban estimates, as well as more accurate disaggregated estimates in slum and non-slum areas. These findings underscore the urgent need for methods to routinely and accurately map deprived urban areas across LMIC countries. Until these maps exist, household surveys based on moderately outdated or inaccurate census data will likely continue to produce biased estimates of urban outcomes, and mask the needs of the urban poorest. These findings also indicate that use of a gridded population sample frame, even if based on an outdated or inaccurate census, produces more accurate estimates than sampling directly from an outdated or inaccurate census.

In Chapter 6 I evaluated gridded population sampling in a real-world setting by partnering with HERD International in Kathmandu, Nepal under the Surveys for Urban Equity project. In this chapter I evaluated the feasibility of gridded population sampling with a standard two-stage sampling design, as well as an area-microcensus design in which all households were listed and interviewed in a small area on the same day. Findings from this survey suggested that area-microcensus sampling is cheaper than two-stage sampling due to smaller PSUs and less time spent mapping and listing households. Results also suggested that when interviewers perform the household listing in area-microcensus PSUs, that a greater number of vulnerable and mobile households can be identified and thus included in the sample. This is likely because interviewers spend much more time with residents in their homes building rapport, while mapping-listing teams only ever have a few minutes at each building. As LMIC cities grow increasingly complex and dynamic, it might be important for survey teams to consider shifting the listing task to interviewers to be able to identify informal and “hidden” households.

I assessed the experiences of field staff in the Kathmandu survey via focus group discussions and shadowing them during fieldwork. Mapping and listing staff generally reported neutral and positive experiences with gridded population survey methods and tools, and strongly preferred the geographically-accurate OpenStreetMap-based maps and tablet navigation applications we used over standard hand-drawn paper maps. However, the field staff suggested hiring non-technical staff to be part of the mapping-listing teams to improve interactions with members of the public, and highlighted challenges that resulted from un-integrated field tools.

7.2 Recommendations and directions for LMIC surveys

The simulated “true” population that I developed in Khomas was highly realistic; the point locations and approximate population densities aligned with actual building locations as observed in satellite imagery collected in the respective year. The household sizes in “poor” and “non-poor” households followed a distribution measured in an actual household survey data, and the location of “poor” and “non-poor” households were realistic based on visual inspection of satellite imagery. Given that this simulated population resembles an actual population, conclusions from this PhD are relevant to all WorldPop-RF / WorldPop-Global data, and to all gridded population surveys in LMICs, particularly in countries with concentrated population surrounded by vast open spaces, such as Argentina, DR Congo, Libya, Pakistan, and Iraq.

Gridded population sampling has potential to improve household survey accuracy in LMICs, particularly if the existing census is moderately outdated (e.g., 10 years) or inaccurate. Improving the accuracy of LMIC household survey data stands to improve the health and wellbeing of the world's poorest by ensuring that they are reflected in monitoring efforts such as the SDGs, and visible in data used by decision-makers to determine and fund local, national, and global development initiatives.

This thesis makes the first attempt to assess cell-level accuracy of a gridded population dataset. I found evidence that WorldPop gridded population sample frames derived from outdated, inaccurate census input data perform better than the original outdated, inaccurate census data in terms of indicator accuracy, filling a critical gap in LMICs. Gridded population sample frames have the advantage over census sample frames of being extremely fine scale (e.g., 100 metres by 100 metres), enabling area-microcensus sample designs. I found evidence that an area-microcensus design used in a complex, dynamic urban setting can better represent certain vulnerable and mobile populations than the standard two-stage design.

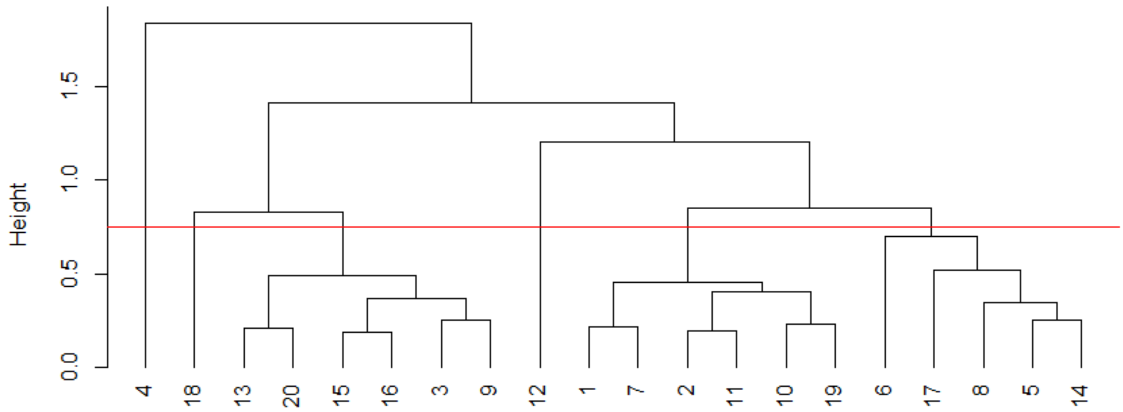
However, two key issues need to be addressed before gridded population sampling can improve survey data accuracy in practice. First, gridded population accuracy needs to be evaluated at the cell-level by the producers of gridded population data to drive research and development of methods for fine-scale gridded population data. In my evaluation of WorldPop-RF / WorldPop-Global data, I found major overestimation of rural and low-dense populations and underestimation of high-dense urban and populations, which is corroborated by my visits to field locations where gridded population estimates are made. I suggest incorporating fine-scale population counts into WorldPop training models to improve urban estimates, reducing misallocation of population to low-density areas. However, there are likely other solutions that could improve cell-level accuracy in gridded population models. I see measurement of cell-level error by the geospatial analysts who produce gridded population datasets as the first step in understanding these errors and innovating to address them.

The second key issue is the inability in most countries to stratify heterogeneous urban populations. Producing maps of deprived areas across LMIC cities would not only vastly improve the accuracy and precision of all household survey estimates, it could make the urban poorest more visible in other data sources such as censuses. The IDEAMAPS initiative, which I have been involved with (Thomson, Kuffer, *et al.*, 2020), and the Million Neighbourhoods Initiative (Brelsford, Martin and Bettencourt, 2017; Brelsford *et al.*, 2018) are two of several collaborations aimed at developing routine maps of deprived urban areas at scale.

Research and practice are moving toward cell-level accuracy assessments of gridded population data and routine mapping of LMIC deprived urban areas, which is promising for gridded population sampling. Gridded population sampling is thus poised to shape the next era in LMIC household surveys.

Appendices

Appendix A Alternative household type clustering results in Khomas, Namibia population simulation



Cluster	urban_ rural	no_edu	any_u5	unimp. toilet	unimp. water	inadeq. structure	limited space	solid fuel	Household type label
Type 1	0	0.02	0.28	0.17	0.00	0.03	0.08	0.01	Urban highest (richest) ??
Type 2	0	0.11	0.34	0.98	0.00	0.55	0.08	0.19	Urban middle-low ??
Type 3	1	0.33	0.40	1.00	0.13	0.53	0.14	1.00	Rural poor
Type 4	0	0.04	0.36	0.70	0.02	0.23	0.16	0.02	Urban middle-high ??
Type 5	1	0.04	0.21	0.43	0.00	0.08	0.00	0.00	Rural rich
Type 6	0	0.07	0.50	0.77	0.07	0.36	0.21	0.79	Urban lowest (poorest)
Khomas	0.06	0.05	0.31	0.50	0.01	0.21	0.09	0.09	

Appendix B Detailed literature review results: Percent population missing from LMIC censuses

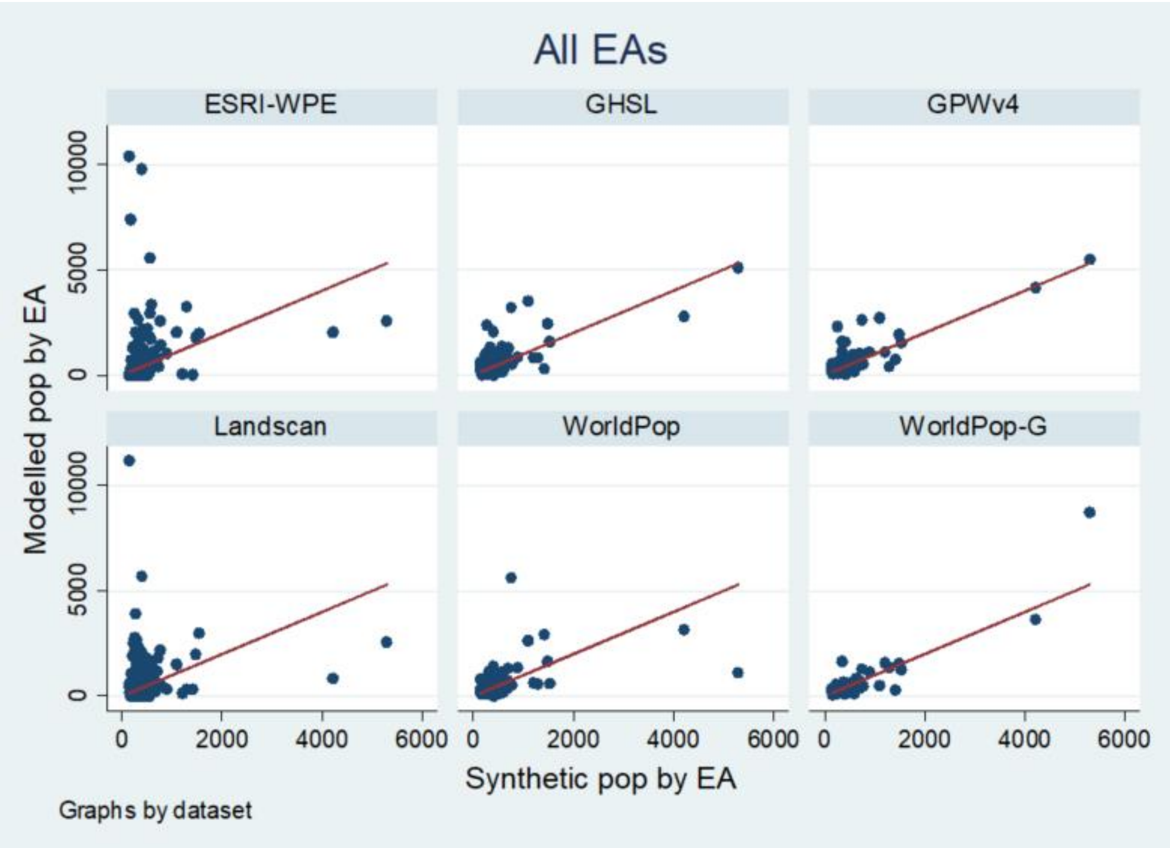
Location & Housing Type	Agarwal (2011) India	Carr-Hill (2013) LMICs	Carr-Hill (2017) Tanzania, Kenya, Uganda	Ebenstein (2015) China	Gidado (2013) Nigeria	Gurgel (2003) Brazil (kids only)	Jiang (2015) China	Karanja (2010) Kenya	Kronefeld (2008) Afghan-istan	Lucci (2018) Kenya	Sabry (2010) Egypt
Urban slum											
Permanent/semi-permanent (not mobile)	50%	5-13%	17-51%					21%		18, 21, 38, 59%	45%
Permanent/semi-permanent (mobile)				33-61%					44%		
Homeless		100%		33-61%		64%					
Urban non-slum											
Permanent							2%				
Rural											
Permanent/semi-permanent							2%				
Remote					12%						
Nomad					12%						
Institutional											
Hospital / care home		<1%									
Prison		0.09									
Refugee camp		10-15%									
Citation	(Agarwal, 2011)	(Carr-Hill, 2013)	(Carr-hill, 2017)	(Ebenstein and Zhao, 2015)	(Gidado *et al.*, 2013)	(Gurgel *et al.*, 2004)	(Jiang *et al.*, 2015)	(Karanja, 2010)	(Kronenfeld, 2008)	(Lucci *et al.*, 2018)	(Sabry, 2010)

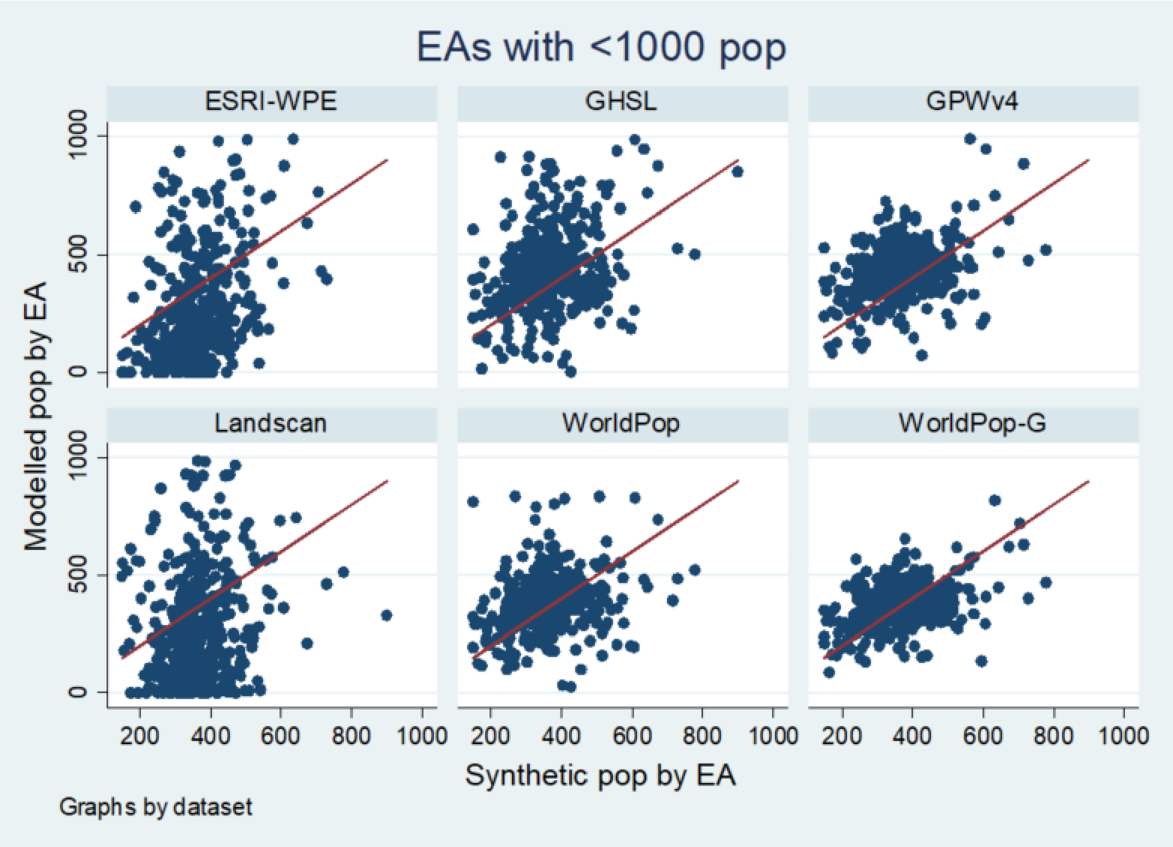
continued…

Location & Housing Type	Stark (2017) Cambodia (kids only)	Treiman (2005) China	PES (2000) Brazil	PES (2001) Nepal	PES (2002) Tanzania	PES (2002) Uganda	PES (2010) Ghana	PES (2010) Zambia	PES (2010) Rwanda	PES (2011) Bangladesh	PES (2011) India	PES (2011) South Africa
Urban slum												
Permanent/semi-permanent (not mobile)												
Permanent/semi-permanent (mobile)		50%										
Homeless	80-96%	50%										
Urban non-slum												
Permanent			2-4%	12%	7%	12%	3%	5%	2%	5%	3%	15%
Rural												
Permanent/semi-permanent			4-11%	5%	7%	5%	2%	10%	2%	4%	2%	13%
Remote												
Nomad												
Institutional												
Hospital / care home												
Prison												
Refugee camp												
Citation	(Stark *et al.*, 2017)	(Treiman *et al.*, 2005)	(Oliveira *et al.*, 2003)	(Korale, 2002)	(Maro, 2009)	(UBS, 2005)	(GSS, 2012)	(CSO, 2013)	(NISR, 2010)	(BIDS, 2012)	(NSC, 2014)	(SSA, 2012)

Appendix C Comparison of existing gridded population datasets with a simulated population in Oshikoto, Namibia

The follow comparison is of a published simulated "realistic" population in the rural region of Oshikoto, Namibia (Thomson, Kools and Jochem, 2018) and all available gridded population datasets on the POPGRID platform (POPGRID Data Collaborative, 2018) including 2016 ESRI World Population Estimates (WPE) (Frye *et al.*, 2018), the 2015 Global Human Settlement Layer (GHSL) Population (also called GHS-POP) (European Commission JRC, 2020), 2015 Global Population of the World v4 (Center for International Earth Science Information Network (CIESIN) and Columbia University, 2016), 2015 LandScan layer (Dobson *et al.*, 2000), and 2014 WorldPop-Random Forest layer (Stevens *et al.*, 2015), plus the 2014 WorldPop-Global dataset (WorldPop, 2019) which I downloaded separately. These comparisons are of the total estimated population per census enumeration area (EA) compared to total households by EA in the simulated dataset, where EA boundaries were defined by the Namibia 2011 census (NSA, 2011a).



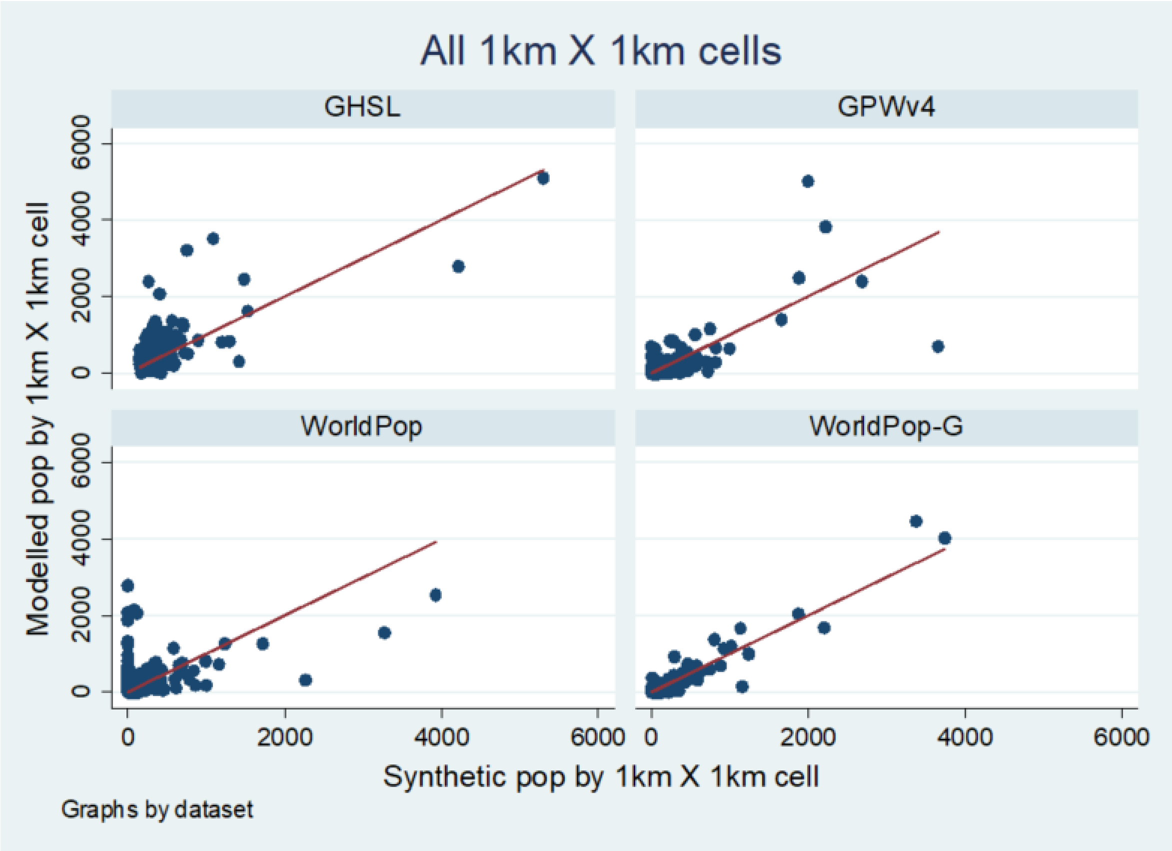


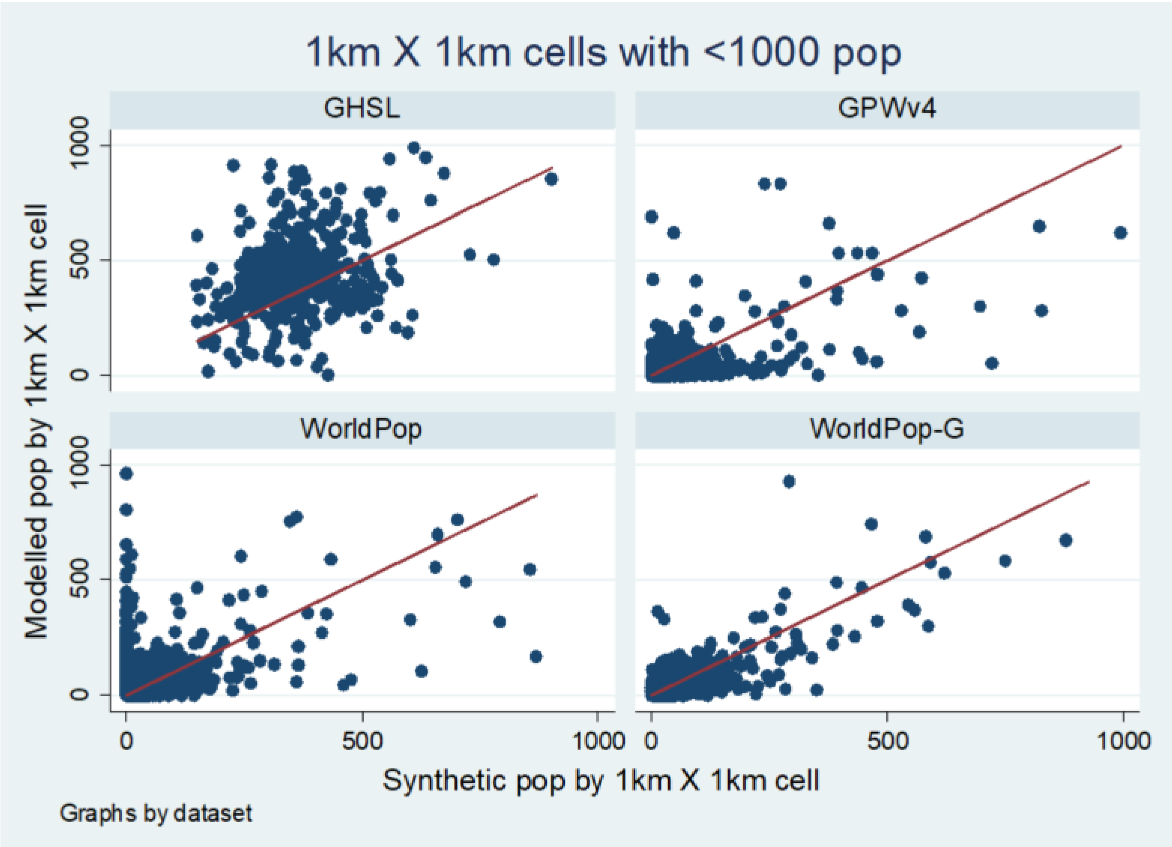
dataset	+/-20	+/-50	class +/-100	+/-1000	>+/-1000
ESRI-WPE	4.01	7.13	7.80	74.61	6.46
GHSL	8.46	12.25	23.16	54.79	1.34
GPWv4	11.36	12.47	26.06	49.00	1.11
Landscan	5.57	6.01	9.80	73.05	5.57
WorldPop	11.14	18.49	30.73	38.53	1.11
WorldPop-G	14.92	20.27	34.52	29.62	0.67
Total	9.24	12.77	22.01	53.27	2.71

dataset	RECODE of class +/-100	>+/-100
ESRI-WPE	18.93	81.07
GHSL	43.88	56.12
GPWv4	49.89	50.11
Landscan	21.38	78.62
WorldPop	60.36	39.64
WorldPop-G	69.71	30.29
Total	44.02	55.98

I observed the greatest consistency in estimates at the EA-level between the simulated population and GPW4, WorldPop-Random Forest, and WorldPop-Global. LandScan and ESRI-WPE performed particularly poorly in this rural area of Namibia.

The following comparisons are of the simulated population and all gridded population estimates that are publicly available, including GHSL, GPWv4, WorldPop-Random Forest, and WorldPop-Global. I downloaded and aggregated each dataset to approximately 1 kilometre by 1 kilometre units; aggregation was performed within each dataset to preserve the estimates and coverage of their original grid cell units. I then joined the simulated household counts to these approximately 1 kilometre by 1 kilometre areas for each dataset to make comparisons.


All 1km X 1km cells
GHSL
GPWv4
WorldPop
WorldPop-G
Modelled pop by 1km X 1km cell
0
2000
4000
6000
Synthetic pop by 1km X 1km cell
Graphs by dataset


1km X 1km cells with <1000 pop
GHSL
GPWv4
WorldPop
WorldPop-G
Modelled pop by 1km X 1km cell
0
500
1000
Synthetic pop by 1km X 1km cell
Graphs by dataset

dataset	class +/-10	+/-20	+/-50	+/-100	>+/-100
GHSL	4.90	3.56	12.25	23.16	56.12
GPWv4	92.21	5.09	2.28	0.28	0.15
WorldPop	89.39	5.25	3.60	1.25	0.50
WorldPop-G	92.46	4.90	2.22	0.30	0.13
Total	90.96	5.09	2.79	0.72	0.44

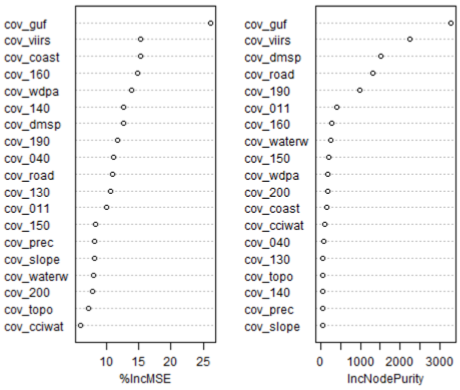
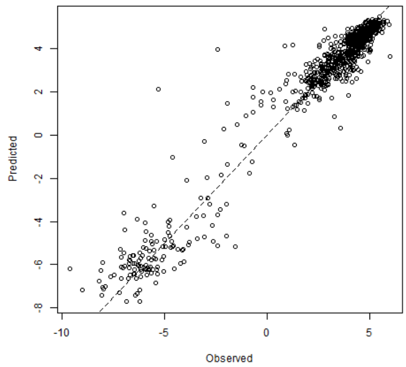
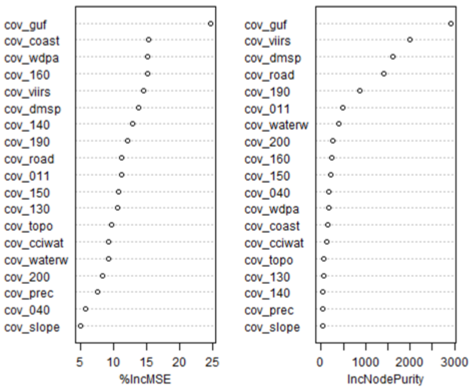
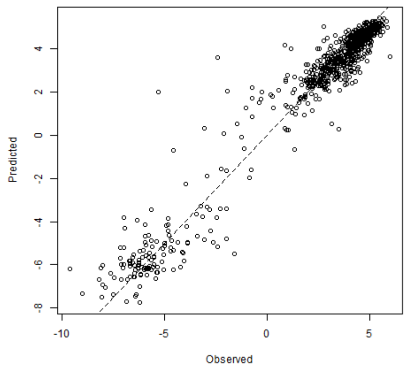
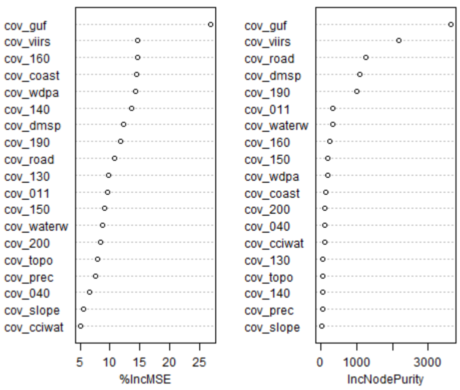
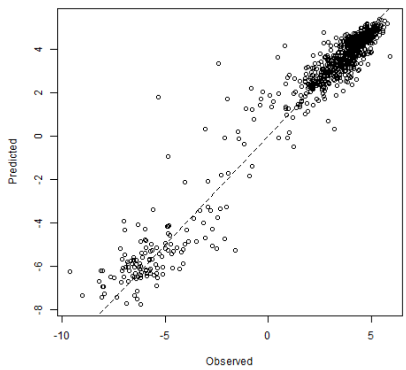
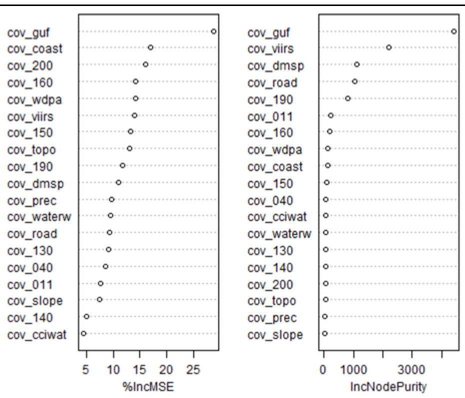
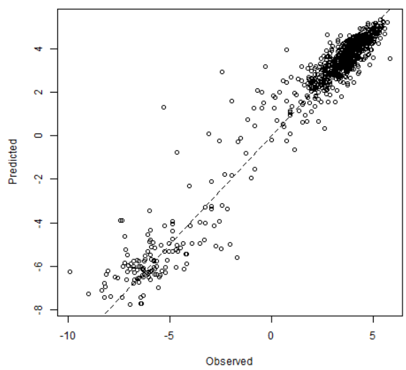
dataset	RECODE of class +/-20	>+/-20
GHSL	8.46	91.54
GPWv4	97.30	2.70
WorldPop	94.64	5.36
WorldPop-G	97.35	2.65
Total	96.05	3.95

At the 1 kilometre by 1 kilometre scale, GPWv4 and WorldPop-Global were most consistent with the simulated population in rural Oshikoto, Namibia, and WorldPop-Random Forest was only slightly less consistent. GHSL population estimates were highly inconsistent.

These findings only demonstrate which gridded datasets perform accurately in a rural contest with a highly dispersed farming population. Datasets which constrain estimates to settled areas (LandScan, ESRI-WPE, and GHSL) are expected to perform poorly in rural areas because the settlement layers they use to constrain population estimates are likely to miss very small settlements.

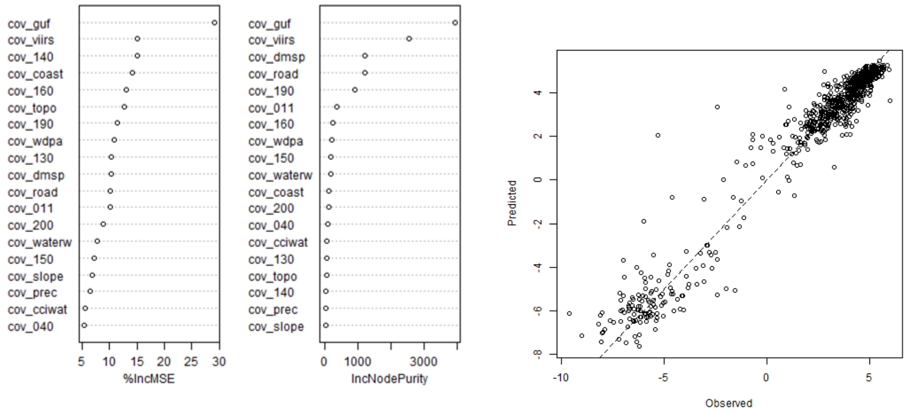
If this same analysis were performed in an area with highly concentrated urban populations, or vast unsettled areas such as deserts or forests, we would expect very different results.

Appendix D Random Forest fit plots for two sets of EA-level models and one set of constituency-level models

Scenario	Model Variance Importance	Observed vs Predicted log(pop)
EA_2016_true Khomas only training data		
EA_2016_L Khomas only training data		
EA_2016_M Khomas only training data		
EA_2016_H Khomas only training data		

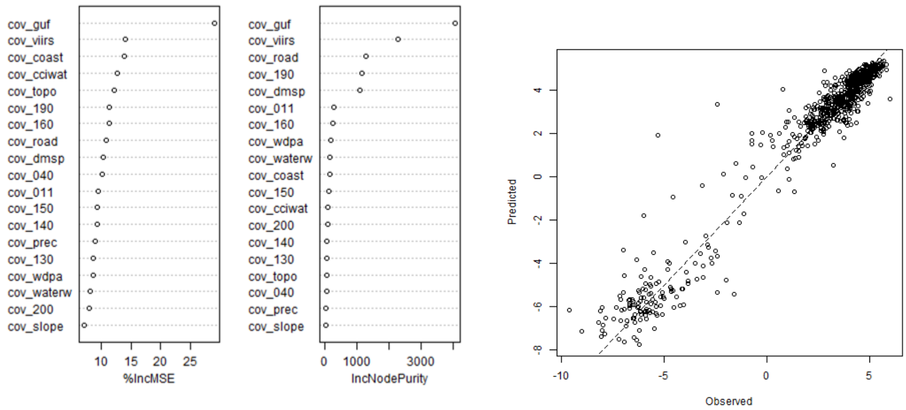
EA_2011_true

Khomas only training data



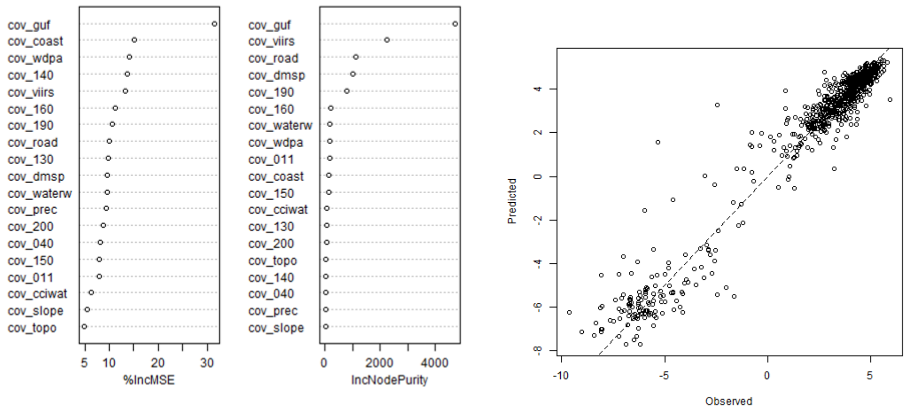
EA_2011_L

Khomas only training data



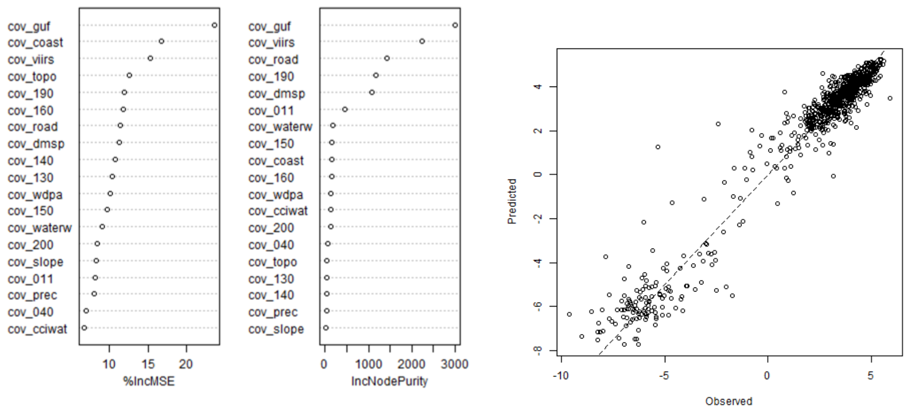
EA_2011_M

Khomas only training data



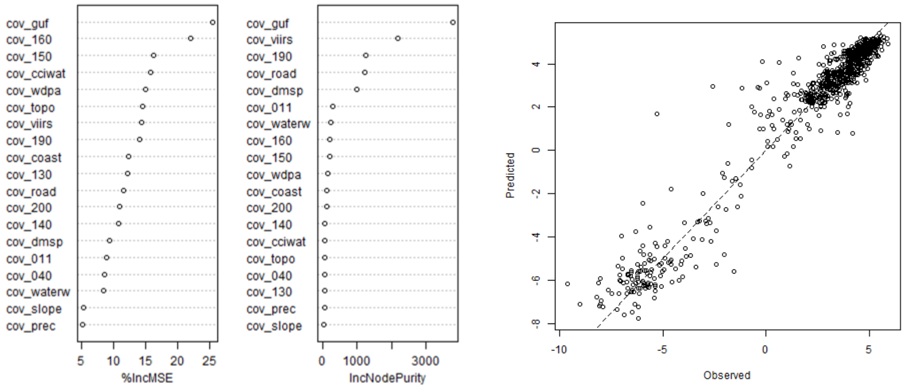
EA_2011_H

Khomas only training data



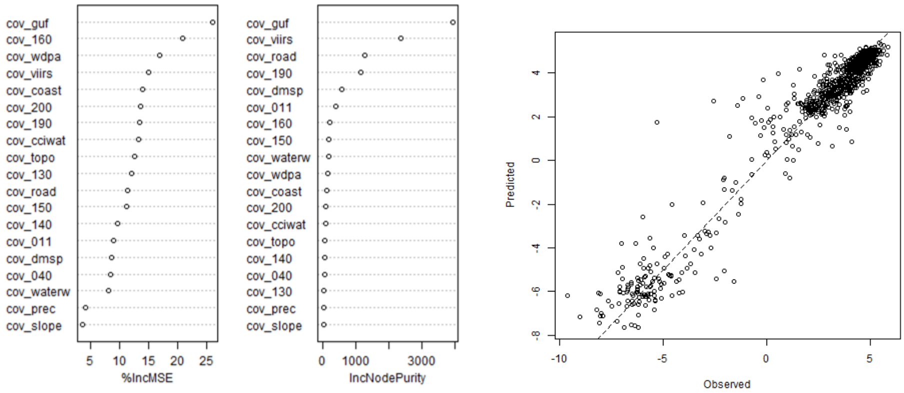
EA_2006_true

Khomas only training data



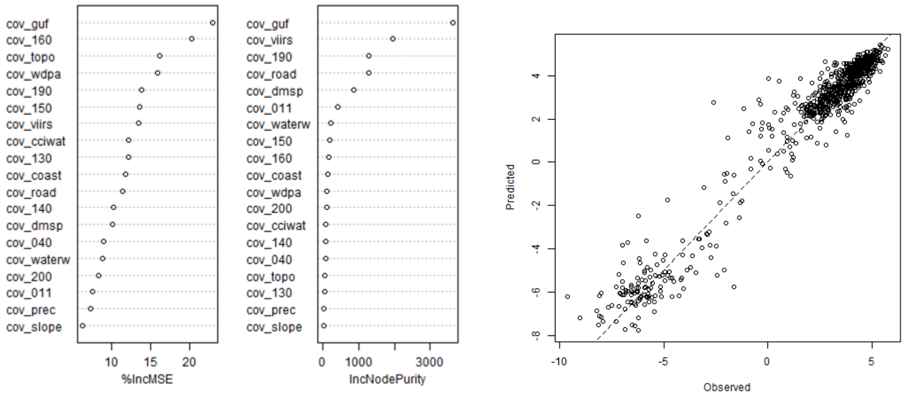
EA_2006_L

Khomas only training data



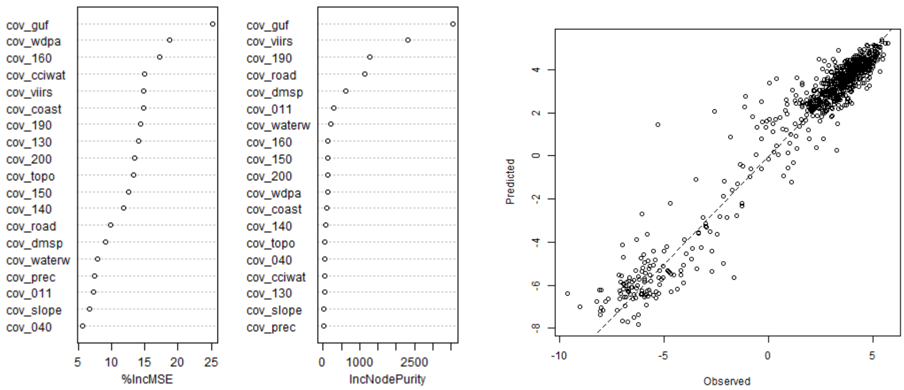
EA_2006_M

Khomas only training data



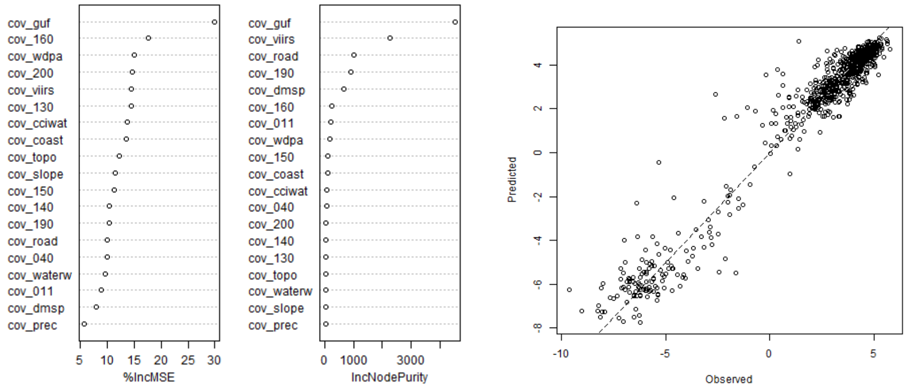
EA_2006_H

Khomas only training data



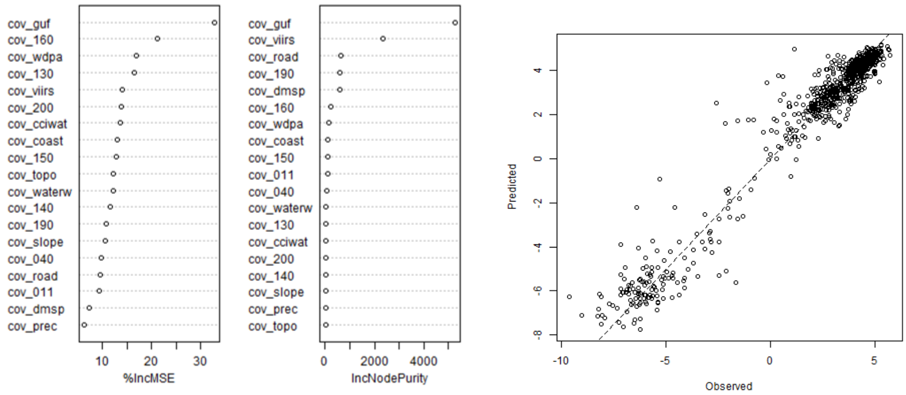
EA_2001_true

Khomas only training data



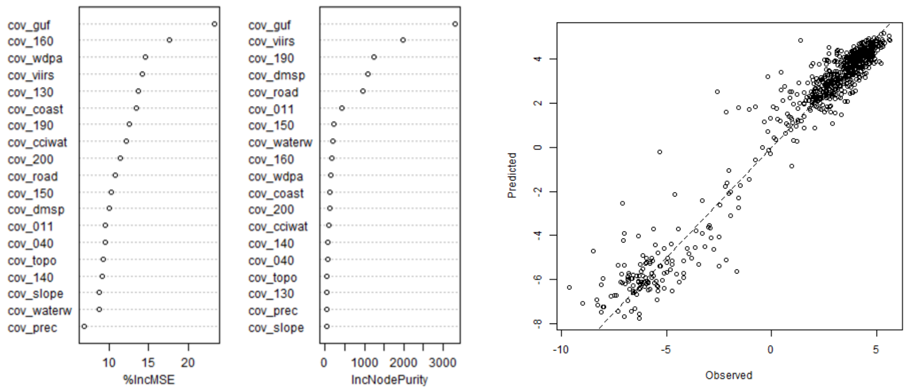
EA_2001_L

Khomas only training data



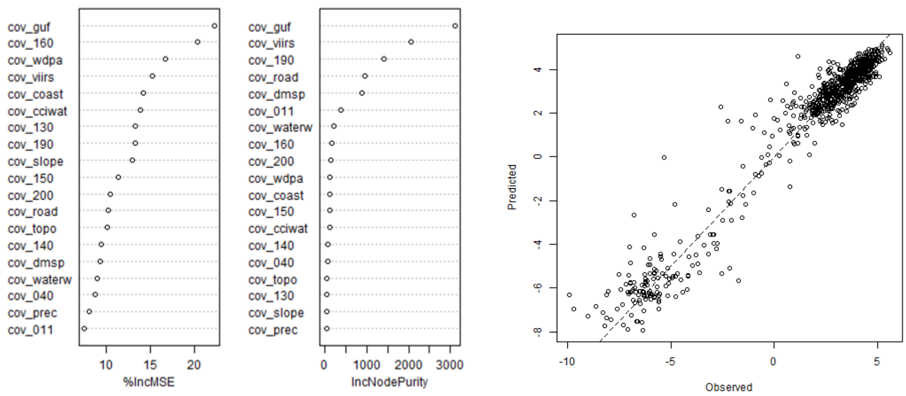
EA_2001_M

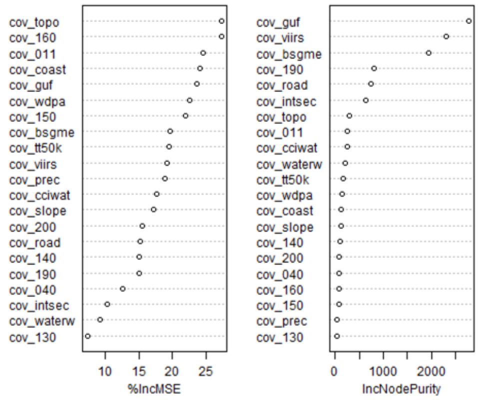
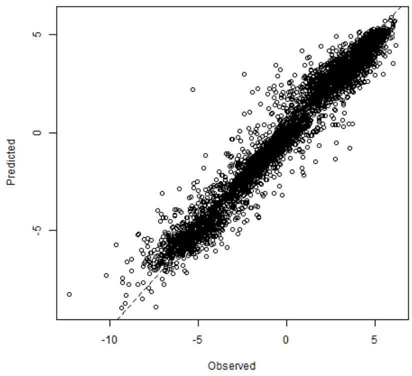
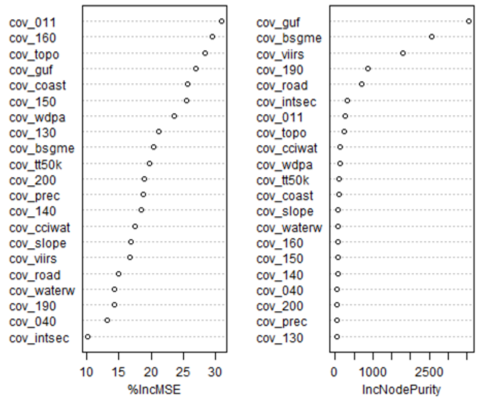
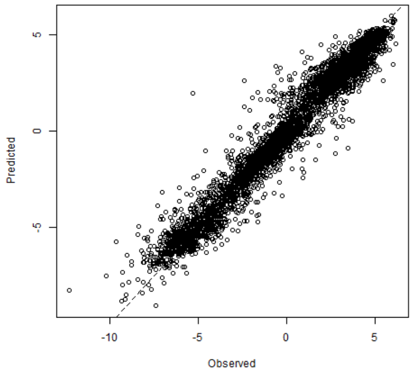
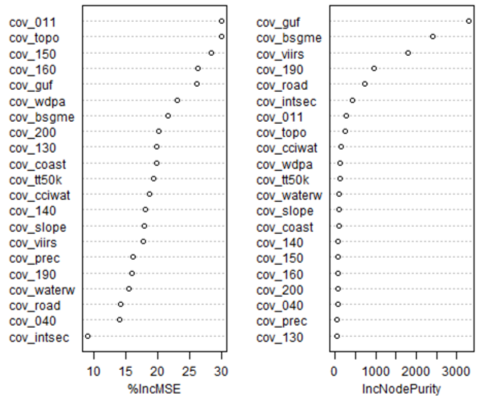
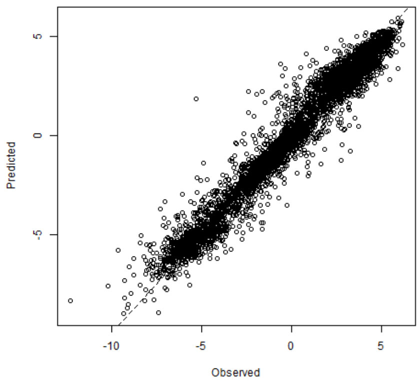
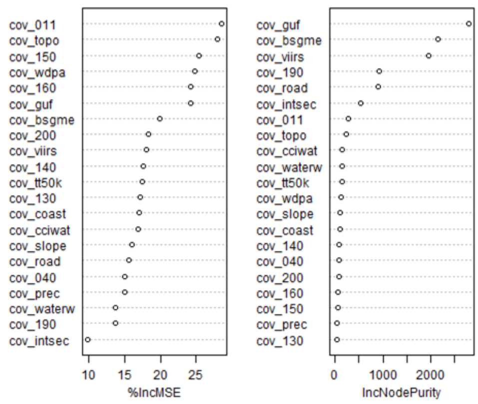
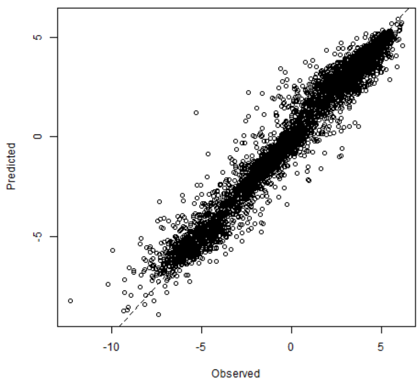
Khomas only training data



EA_2001_H

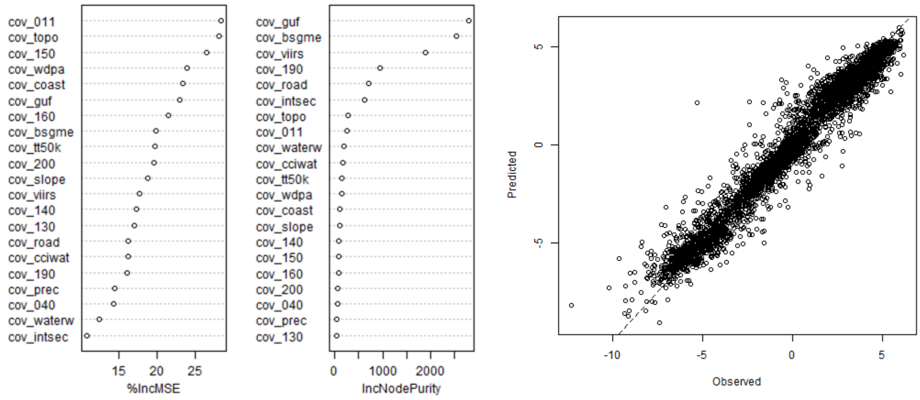
Khomas only training data



Scenario	Model Variance Importance	Observed vs Predicted log(pop)
EA_2016_true Includes training data outside of Khomas		
EA_2016_L Includes training data outside of Khomas		
EA_2016_M Includes training data outside of Khomas		
EA_2016_H Includes training data outside of Khomas		

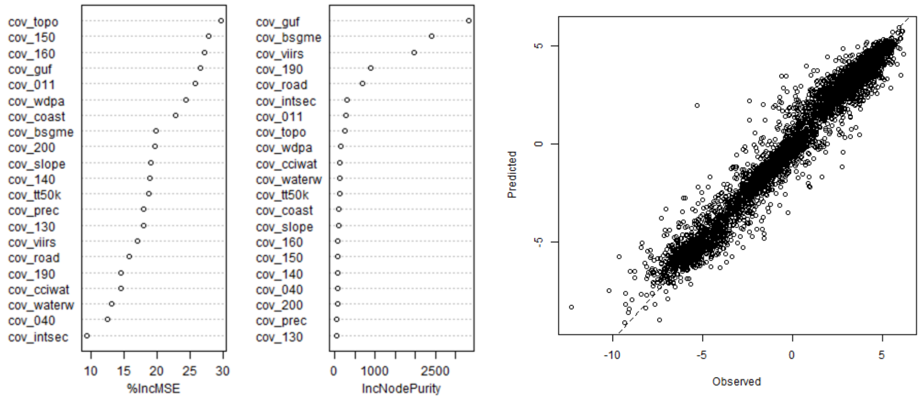
EA_2011_true

Includes training data outside of Khomas



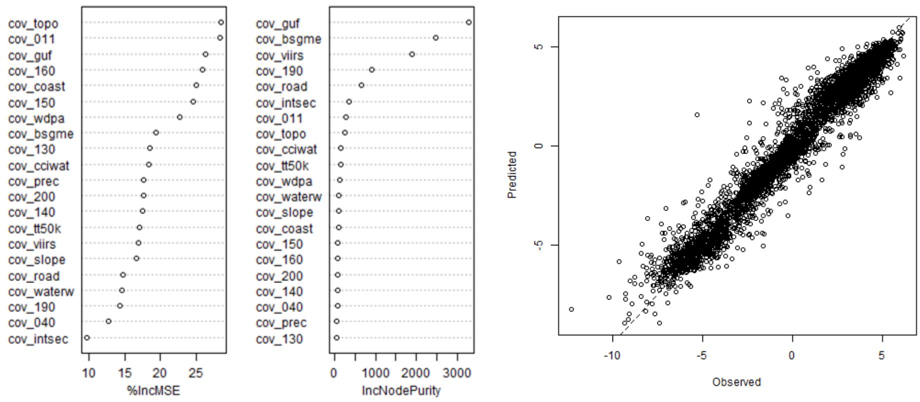
EA_2011_L

Includes training data outside of Khomas



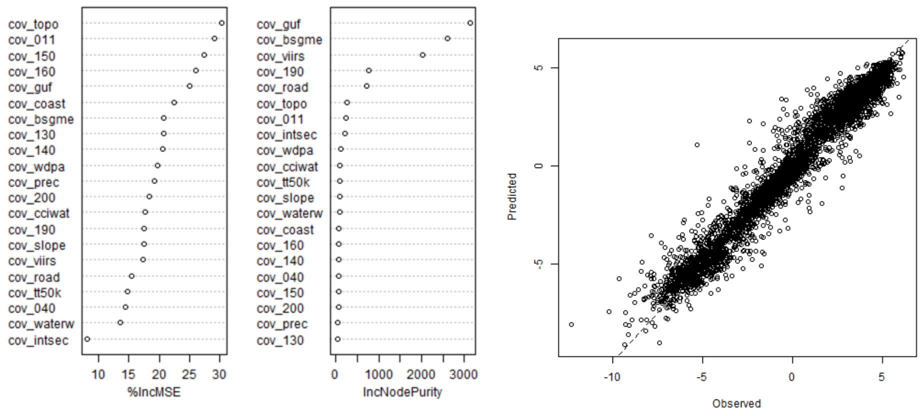
EA_2011_M

Includes training data outside of Khomas



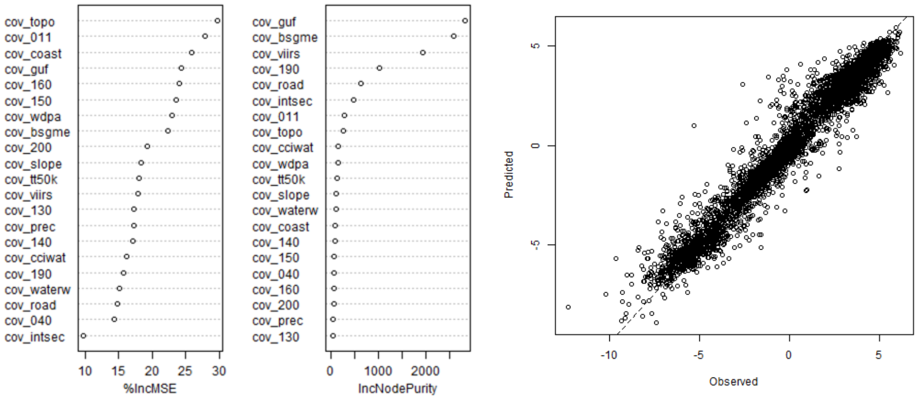
EA_2011_H

Includes training data outside of Khomas



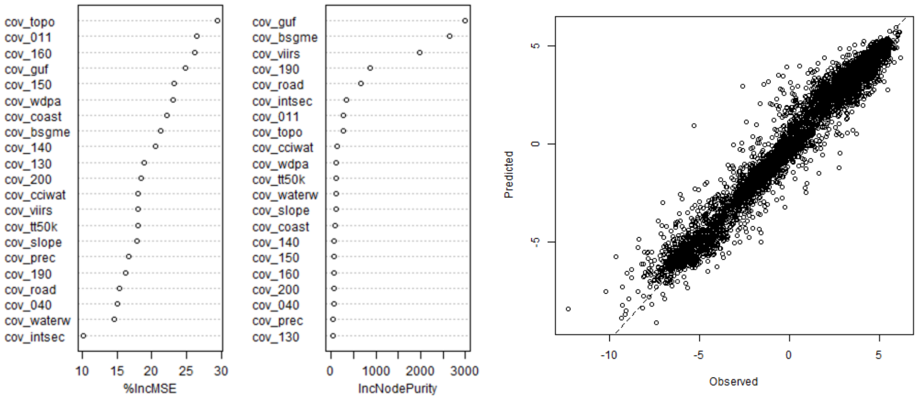
EA_2006_true

Includes training data outside of Khomas



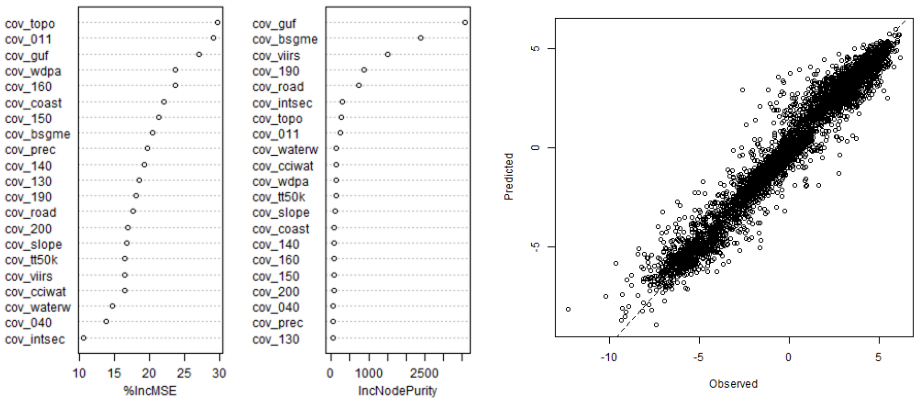
EA_2006_L

Includes training data outside of Khomas



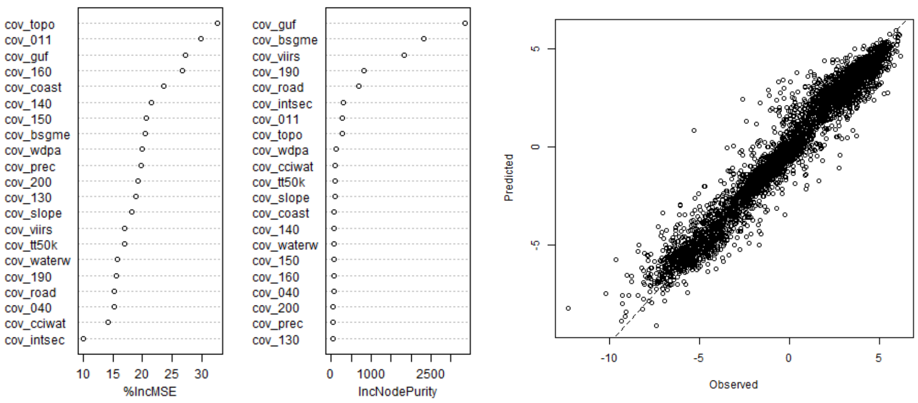
EA_2006_M

Includes training data outside of Khomas



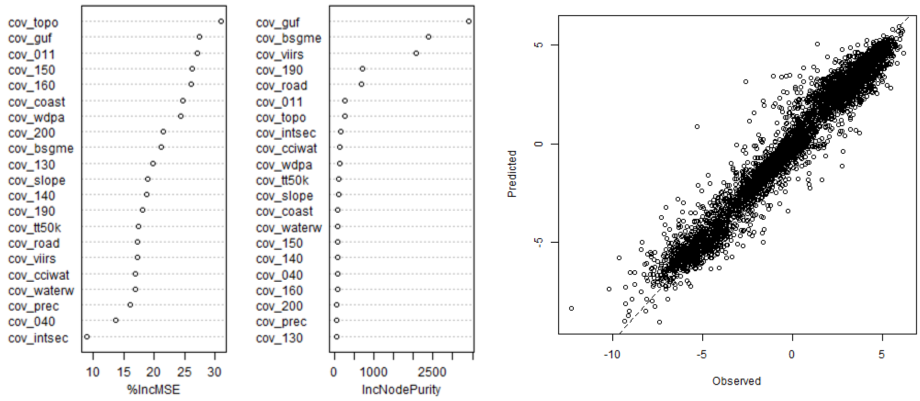
EA_2006_H

Includes training data outside of Khomas



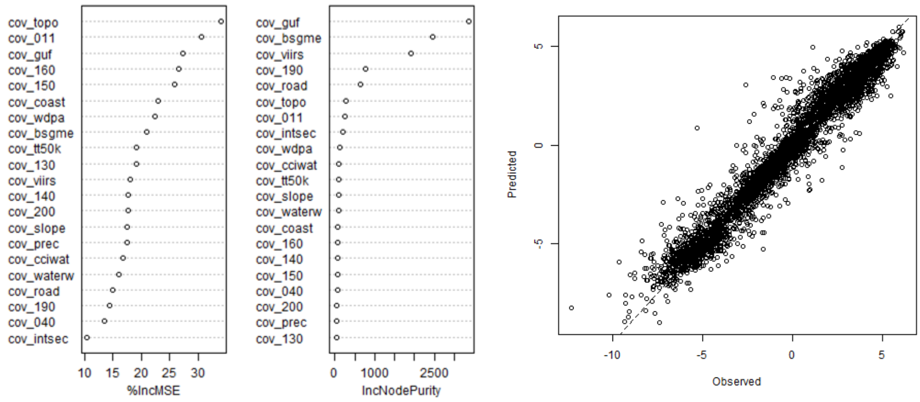
EA_2001_true

Includes training data outside of Khomas



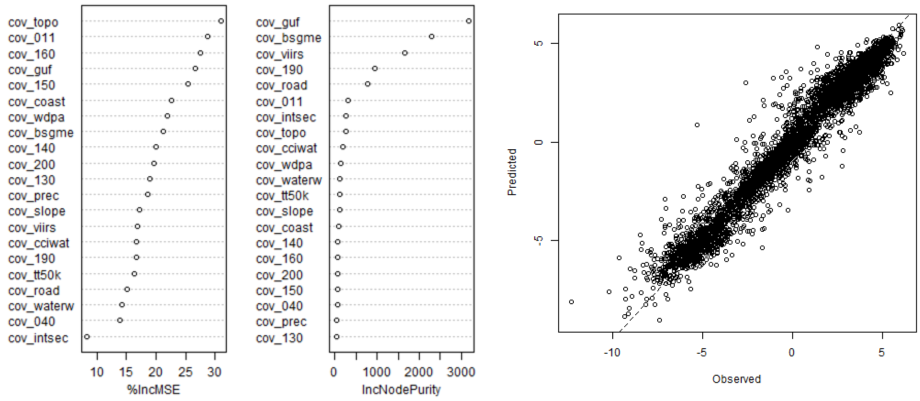
EA_2001_L

Includes training data outside of Khomas



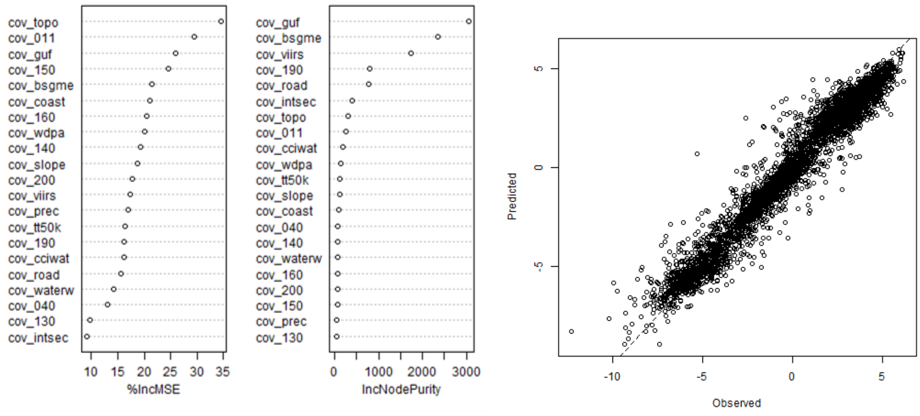
EA_2001_M

Includes training data outside of Khomas



EA_2001_H

Includes training data outside of Khomas

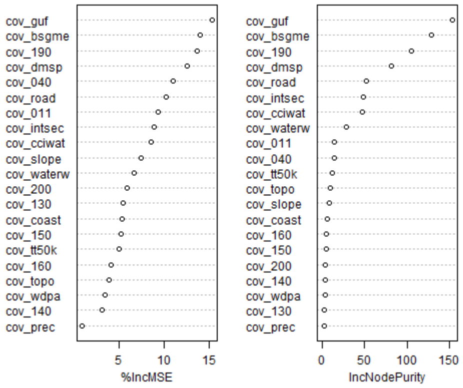


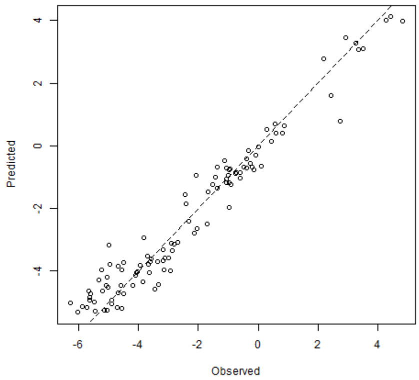
Scenario	Model Variance Importance	Observed vs Predicted log(pop)

Constituency

2016_true

Includes training data outside of Khomas

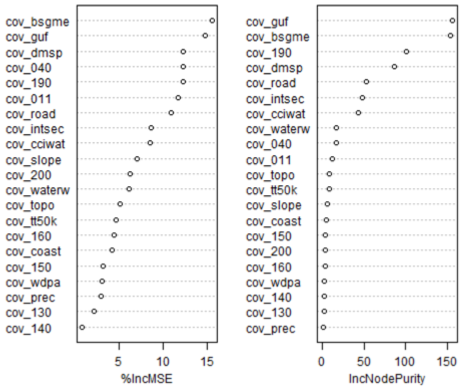


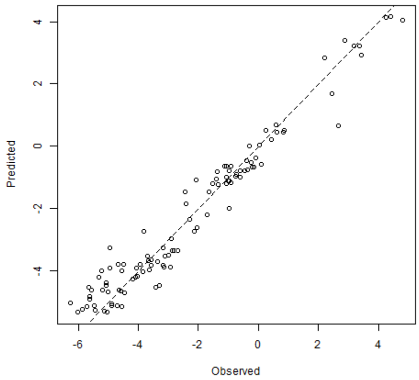


Constituency

2016_L

Includes training data outside of Khomas

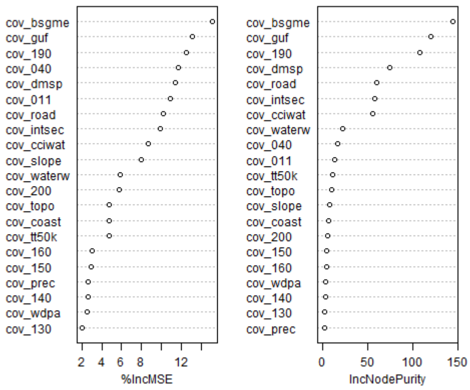


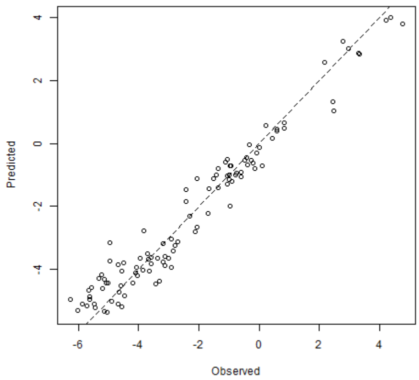


Constituency

2016_M

Includes training data outside of Khomas

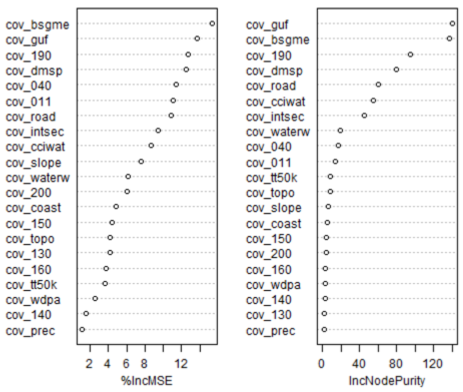


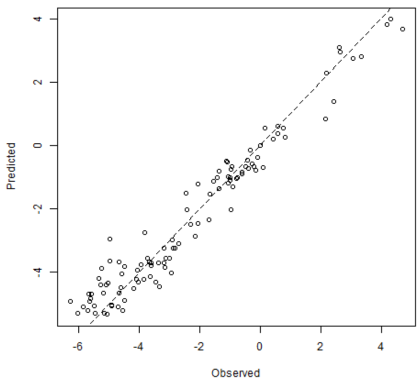


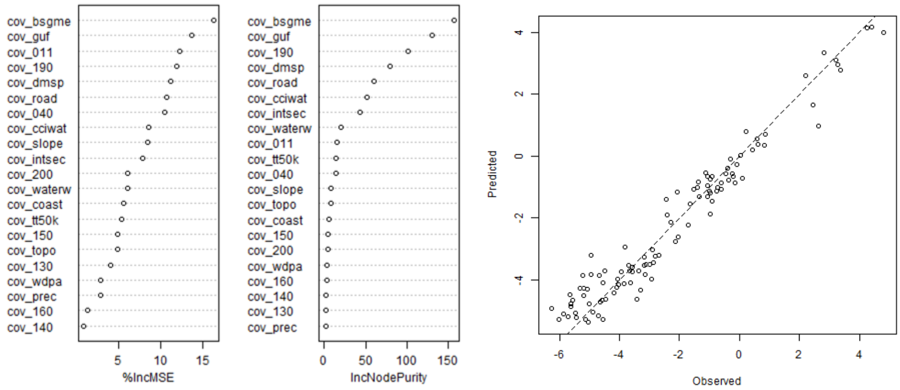
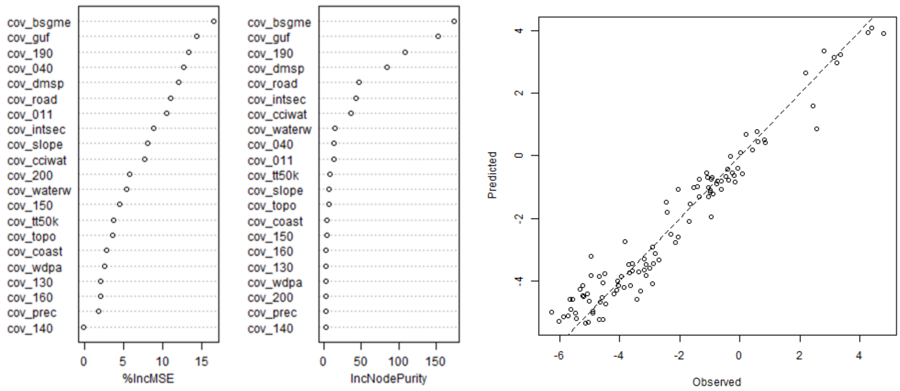
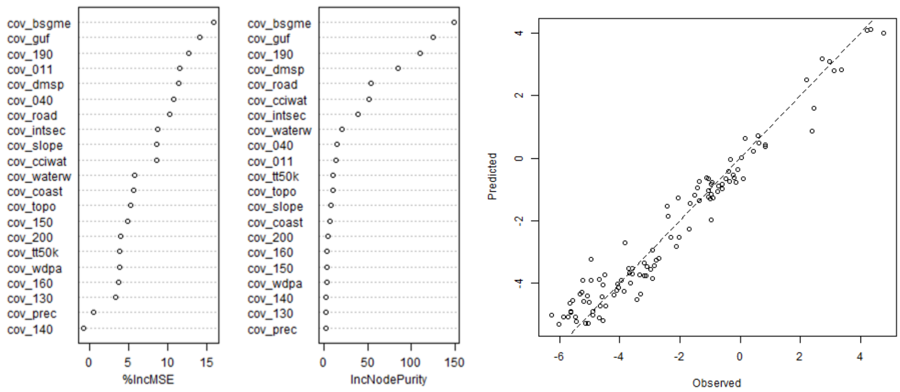
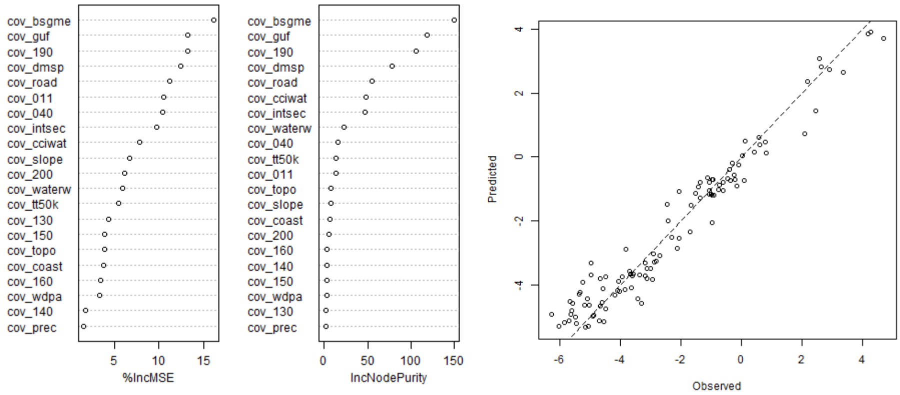
Constituency

2016_H

Includes training data outside of Khomas



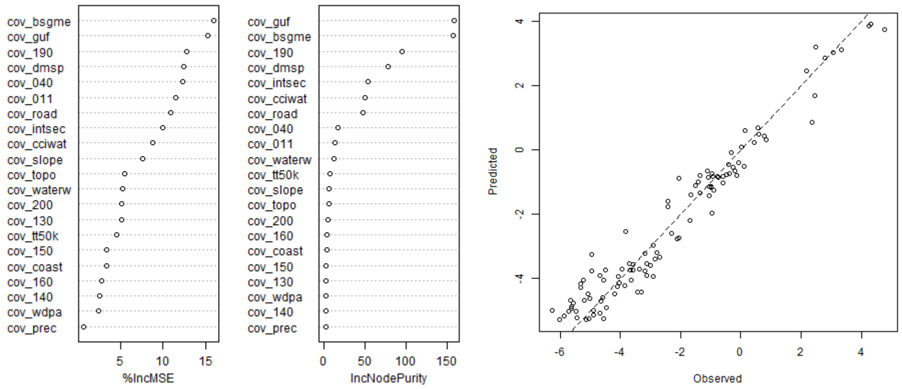


Constituency 2011_true Includes training data outside of Khomas	
Constituency 2011_L Includes training data outside of Khomas	
Constituency 2011_M Includes training data outside of Khomas	
Constituency 2011_H Includes training data outside of Khomas	

Constituency

2006_true

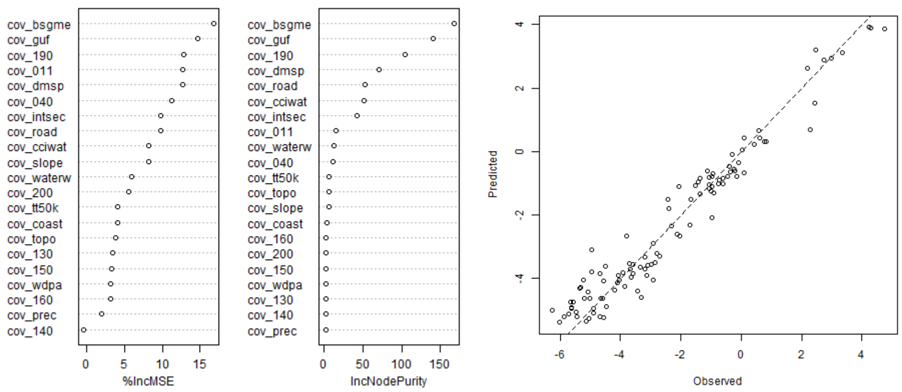
Includes training data outside of Khomas



Constituency

2006_L

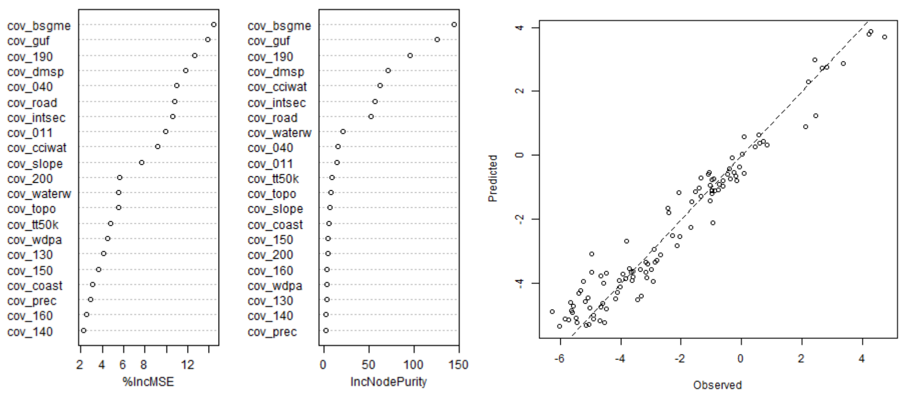
Includes training data outside of Khomas



Constituency

2006_M

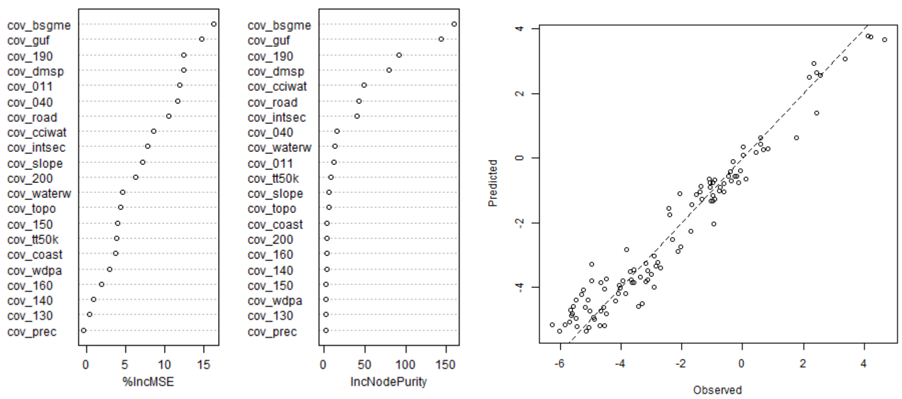
Includes training data outside of Khomas



Constituency

2006_H

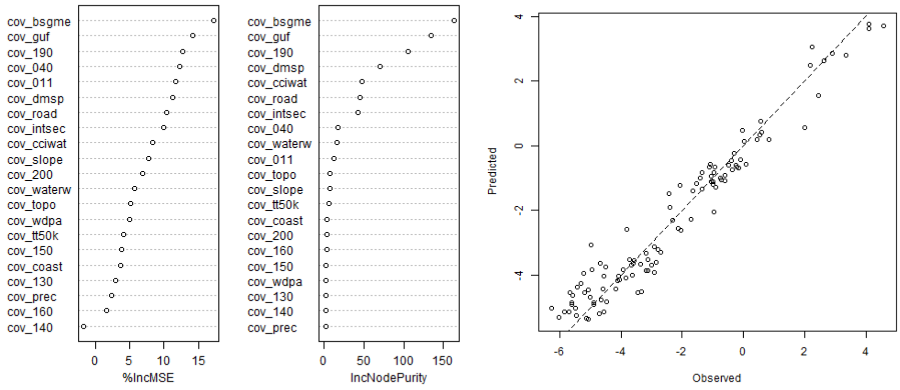
Includes training data outside of Khomas



Constituency

2001_true

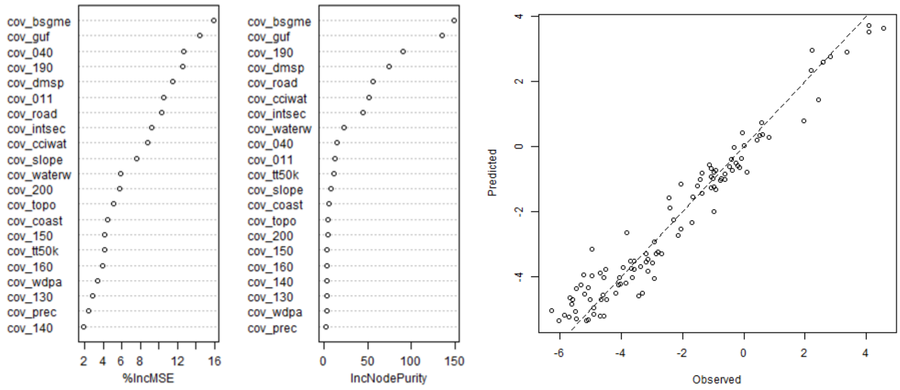
Includes training data outside of Khomas



Constituency

2001_L

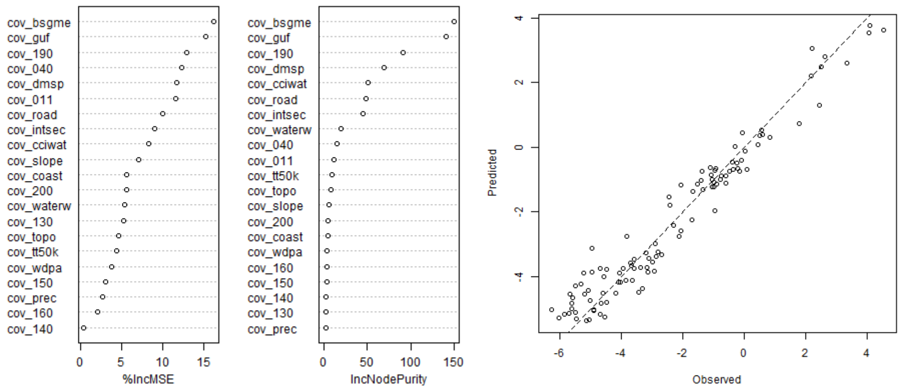
Includes training data outside of Khomas



Constituency

2001_M

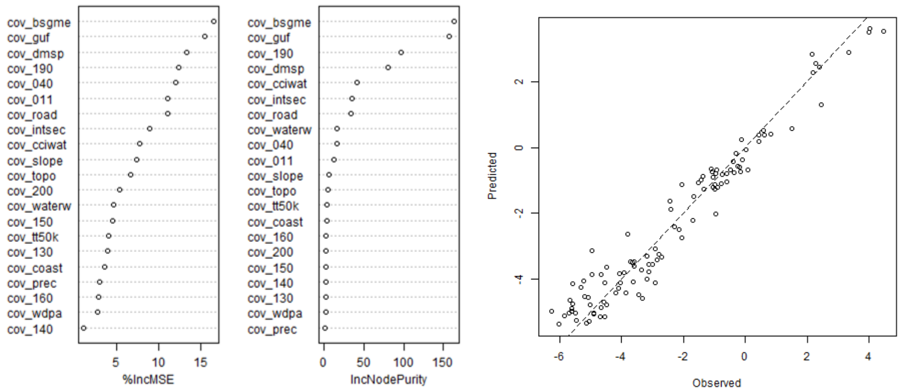
Includes training data outside of Khomas



Constituency

2001_H

Includes training data outside of Khomas



Appendix E Simulated model estimates versus "true" simulated population in Khomas, Namibia

RMSE, EA-level input data, excluding cells with estimated population <1																
Cell size (metres)	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
	All															
100	27	29	26	29	28	30	33	30	31	31	29	31	31	30	30	31
200	99	101	103	109	102	104	108	111	102	103	105	110	105	106	108	111
300	238	240	244	250	243	245	249	254	241	242	245	250	245	246	249	254
400	429	431	435	443	437	439	442	450	434	436	439	447	440	442	445	451
500	660	664	669	680	673	674	678	689	668	670	676	682	673	675	680	684
600	936	939	942	956	951	957	959	970	945	949	952	967	958	959	967	976
700	1258	1258	1267	1282	1282	1283	1289	1303	1272	1275	1281	1295	1297	1301	1303	1307
800	1617	1618	1622	1647	1638	1644	1649	1666	1666	1675	1685	1687	1682	1684	1692	1699
900	1965	1971	1984	2011	1997	2001	2008	2036	2005	2014	2025	2036	2040	2041	2052	2063
1000	2398	2403	2433	2460	2437	2443	2454	2474	2462	2467	2470	2477	2511	2512	2513	2528
	Rural															
100	29	33	27	30	31	34	38	30	35	34	30	31	33	32	30	31
200	30	33	27	30	31	34	38	31	35	34	30	32	33	32	30	31
300	36	39	34	37	38	40	43	38	39	39	35	37	37	36	35	36
400	51	53	50	52	52	54	56	53	53	53	51	52	51	50	50	50
500	70	72	70	72	70	72	74	71	64	63	62	61	57	57	56	55
600	152	154	153	156	152	153	154	154	138	138	138	139	130	130	131	132
700	179	180	180	182	175	176	178	178	163	164	164	165	147	147	147	147
800	278	278	279	283	280	282	283	284	264	265	265	266	260	260	260	262
900	419	421	425	428	424	423	426	431	366	368	368	370	305	305	305	305
1000	525	527	536	542	535	537	542	544	541	541	542	542	536	536	536	539
	Urban															
100	24	24	25	29	24	24	25	30	25	25	27	30	26	27	29	32
200	142	145	150	158	148	150	154	162	151	153	157	164	158	159	163	169
300	353	357	362	373	362	365	370	381	366	369	374	384	377	379	386	394
400	644	648	653	666	658	659	665	679	666	670	675	690	684	687	695	705
500	992	998	1005	1022	1009	1012	1020	1035	1020	1023	1036	1048	1039	1043	1060	1066
600	1423	1428	1432	1459	1442	1452	1456	1477	1457	1464	1471	1496	1493	1494	1513	1526
700	1946	1948	1952	1982	1981	1983	1986	2005	1987	1993	2000	2020	2041	2047	2059	2065
800	2513	2515	2519	2558	2540	2542	2558	2578	2593	2617	2635	2657	2655	2656	2689	2721
900	3048	3049	3063	3109	3089	3091	3094	3141	3135	3147	3173	3200	3227	3228	3255	3296
1000	3718	3720	3738	3773	3763	3765	3773	3824	3811	3829	3844	3865	3931	3932	3934	3973

RMSE, Constituency-level input data, excluding cells with estimated population <1																
Cell size (metres)	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
	All															
100	25	25	25	27	25	25	26	27	26	27	27	29	28	29	29	30
200	111	112	114	117	112	113	115	119	117	118	118	124	123	123	125	129
300	253	255	258	263	255	258	261	267	263	265	266	275	275	273	278	286
400	449	453	456	464	453	457	461	470	464	467	468	485	483	481	487	502
500	689	695	699	710	693	700	707	719	709	714	717	741	737	736	745	768
600	982	991	993	1011	989	993	1002	1023	1007	1014	1015	1048	1045	1039	1052	1089
700	1329	1341	1346	1362	1340	1352	1356	1376	1363	1368	1370	1407	1406	1402	1421	1468
800	1701	1708	1717	1756	1702	1727	1746	1777	1752	1769	1777	1820	1817	1810	1825	1885
900	2087	2112	2120	2164	2106	2117	2144	2186	2154	2174	2179	2249	2234	2227	2239	2310
1000	2581	2615	2598	2648	2592	2612	2642	2685	2648	2662	2673	2761	2738	2739	2773	2846
	Rural															
100	11	10	9	7	9	9	8	6	7	7	6	5	6	6	5	4
200	13	13	12	11	12	12	11	11	11	11	11	11	11	11	11	11
300	25	25	25	25	25	25	25	26	26	26	26	27	27	27	28	29
400	44	44	44	46	44	45	45	47	46	46	46	49	49	48	50	53
500	66	68	68	70	67	68	69	72	70	71	71	75	75	74	76	81
600	158	161	161	166	160	161	164	170	165	167	167	177	176	174	179	192
700	194	198	198	203	197	199	201	207	202	203	203	212	214	213	217	232
800	297	303	302	315	300	305	311	322	313	318	319	335	333	331	337	360
900	453	465	467	485	461	465	477	494	480	490	492	523	515	512	517	554
1000	590	605	597	620	594	603	617	637	620	625	627	670	658	658	676	715
	Urban															
100	33	33	35	37	34	35	35	37	36	37	37	39	38	39	39	40
200	157	158	161	164	159	160	162	164	163	164	165	168	168	169	169	171
300	361	361	366	369	364	365	367	370	369	370	372	376	376	377	377	379
400	642	641	649	653	647	649	651	654	653	654	657	663	662	663	663	665
500	982	983	992	995	986	991	994	996	994	998	1004	1009	1009	1009	1010	1012
600	1409	1410	1418	1426	1413	1417	1421	1429	1428	1430	1431	1439	1439	1442	1440	1444
700	1917	1914	1930	1933	1925	1930	1931	1937	1939	1944	1953	1951	1954	1955	1955	1961
800	2447	2448	2462	2470	2460	2461	2468	2471	2469	2482	2496	2498	2497	2498	2498	2506
900	3031	3022	3034	3045	3033	3034	3035	3046	3045	3045	3047	3058	3058	3058	3059	3061
1000	3702	3702	3705	3708	3703	3704	3706	3708	3707	3719	3735	3737	3736	3737	3738	3739

MAE, EA-level input data, excluding cells with estimated population <1

Cell size (metres)	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
All																
100	17.1	17.9	16.4	18.7	18.3	19.4	21.8	19.5	20.1	19.9	18.9	20.2	20.0	19.7	19.8	20.3
200	50.9	52.4	51.9	55.6	53.3	54.9	58.3	57.0	53.5	53.8	53.6	55.8	53.7	53.6	54.3	55.4
300	112.9	114.6	114.8	119.5	116.3	118.1	121.8	121.7	113.9	114.3	114.2	117.5	114.1	114.0	115.5	117.6
400	198.1	200.4	200.9	206.8	203.8	205.6	208.9	210.5	199.8	200.7	200.8	205.9	199.7	200.4	201.4	204.6
500	304.3	307.4	309.0	317.7	312.1	313.3	317.3	321.2	306.8	307.3	309.1	311.7	302.5	302.9	305.0	306.5
600	428.7	431.8	431.1	442.9	440.0	444.6	447.0	450.9	429.5	431.4	432.0	442.2	429.0	428.7	434.0	438.2
700	571.6	572.3	576.6	588.0	588.1	589.4	595.2	601.2	573.1	574.9	578.0	585.7	576.1	578.2	577.6	579.1
800	738.1	739.9	740.5	759.1	751.4	757.5	760.0	770.9	754.9	760.7	765.5	763.8	752.1	753.2	755.3	756.5
900	911.2	916.9	925.1	944.5	935.6	937.8	944.7	964.4	912.4	919.2	924.4	932.0	917.3	917.1	922.6	927.7
1000	1092.9	1097.5	1119.3	1140.5	1123.0	1127.4	1133.6	1143.8	1121.4	1124.1	1120.0	1123.0	1131.1	1130.9	1131.1	1138.9
Rural																
100	19.8	21.1	17.8	20.1	21.4	23.4	27.1	21.0	24.0	23.6	21.3	22.2	23.3	22.6	22.1	21.8
200	19.9	21.2	18.0	20.2	21.4	23.3	27.0	21.1	23.9	23.5	21.3	22.2	23.2	22.4	22.0	21.7
300	22.1	23.5	20.5	22.5	23.7	25.5	29.1	23.4	25.4	24.9	22.7	23.7	24.4	23.6	23.2	23.0
400	25.8	27.1	24.2	26.3	27.4	29.3	32.6	27.1	28.9	28.6	26.5	27.2	27.5	26.7	26.2	25.8
500	30.6	32.4	29.1	31.3	31.6	33.3	36.6	31.1	31.4	31.1	29.0	29.1	28.6	27.8	27.4	26.0
600	54.0	55.7	52.6	55.4	54.6	56.3	59.8	54.3	49.5	49.2	47.1	47.5	43.9	43.0	42.6	42.3
700	63.9	65.7	63.1	65.1	63.2	64.6	68.2	63.3	60.0	59.8	58.3	59.0	50.1	49.6	48.9	47.9
800	103.3	105.3	102.8	104.4	104.2	106.4	109.0	103.4	93.8	93.8	91.2	91.9	86.9	86.5	85.5	85.1
900	156.1	158.3	156.9	156.7	155.9	155.3	158.2	155.9	128.2	128.8	126.1	127.3	103.4	102.7	101.5	101.2
1000	154.7	156.3	156.1	157.5	158.1	159.5	163.8	159.1	145.5	145.2	143.3	144.0	130.1	129.4	129.0	129.2
Urban																
100	14.2	14.2	14.8	17.0	14.6	14.7	15.5	17.8	14.8	15.0	15.7	17.5	15.4	15.7	16.5	18.2
200	87.2	89.1	92.0	97.7	91.5	93.0	95.8	100.9	92.5	93.8	96.2	101.1	95.6	96.8	99.6	103.0
300	223.6	226.5	230.4	239.5	231.3	233.0	236.8	245.6	232.4	234.1	237.9	246.6	238.7	239.6	246.3	252.5
400	416.3	419.8	423.5	434.6	427.9	428.5	432.8	446.6	432.8	435.7	440.3	453.9	444.7	446.9	454.9	463.3
500	650.1	656.5	662.8	679.5	664.8	667.3	674.9	689.2	674.4	676.4	689.2	698.3	681.2	685.2	703.0	707.1
600	933.4	937.4	939.6	969.3	951.1	962.5	964.0	986.5	961.9	967.9	975.5	1001.2	986.8	987.6	1008.0	1019.2
700	1293.5	1294.8	1297.1	1330.8	1329.7	1330.8	1333.0	1351.0	1323.8	1329.5	1336.3	1353.9	1361.5	1366.4	1379.0	1382.8
800	1663.9	1665.1	1667.5	1711.4	1687.9	1689.0	1702.6	1728.7	1717.5	1744.9	1765.0	1783.6	1767.9	1767.2	1802.5	1832.1
900	2025.8	2027.0	2041.9	2093.5	2075.5	2076.6	2078.5	2134.7	2085.7	2099.1	2127.9	2157.5	2167.9	2168.6	2197.8	2241.5
1000	2476.2	2477.3	2498.3	2538.5	2525.9	2526.9	2523.6	2582.2	2552.4	2572.0	2578.1	2597.2	2651.9	2652.6	2653.8	2692.6

MAE, Constituency-level input data, excluding cells with estimated population <1

Cell size (metres)	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
All																
100	13.9	13.8	13.7	13.6	13.7	13.8	13.7	13.8	13.9	13.8	13.7	14.4	14.6	14.4	14.5	15.0
200	49.4	49.9	50.2	51.5	49.7	50.5	51.0	52.8	51.9	52.4	52.3	55.9	56.0	55.5	56.7	59.8
300	111.6	113.0	113.8	117.4	112.4	114.2	116.3	120.3	117.6	118.9	119.0	126.3	126.7	125.1	128.4	135.8
400	200.2	203.5	204.1	210.3	202.6	205.7	208.5	215.1	210.7	212.8	213.3	227.0	226.2	224.0	229.3	242.6
500	312.5	317.4	319.4	327.9	315.4	320.6	325.6	335.5	327.7	331.2	333.3	354.4	351.6	349.9	358.9	379.7
600	450.9	458.8	458.7	473.3	456.0	459.3	466.4	484.9	470.8	476.1	476.4	506.3	504.1	498.4	510.5	545.9
700	616.6	626.9	629.1	642.9	624.8	634.8	638.7	655.4	643.9	647.8	649.4	682.8	683.8	679.9	697.2	742.9
800	791.8	799.0	804.5	838.6	792.8	814.0	829.8	857.8	834.4	851.2	856.5	897.7	895.4	888.6	903.0	961.7
900	998.0	1023.6	1028.9	1068.3	1016.3	1025.7	1051.4	1089.0	1058.4	1078.6	1082.5	1152.3	1137.1	1129.9	1141.4	1214.5
1000	1236.8	1268.1	1250.4	1298.7	1244.5	1264.0	1292.2	1333.3	1298.5	1308.0	1317.6	1405.0	1383.1	1381.5	1417.1	1490.6
Rural																
100	7.5	7.0	6.2	5.0	6.8	6.5	5.7	4.7	5.4	5.1	4.5	3.9	4.7	4.4	4.0	3.4
200	7.9	7.4	6.7	5.6	7.2	7.0	6.3	5.5	6.1	5.8	5.2	4.9	5.6	5.2	4.9	4.6
300	10.0	9.7	9.1	8.4	9.5	9.4	8.9	8.4	8.8	8.6	8.1	8.2	8.8	8.3	8.3	8.5
400	13.8	13.8	13.1	12.7	13.4	13.5	13.0	12.9	13.0	12.9	12.5	13.2	13.6	13.2	13.4	14.2
500	19.1	19.1	18.6	18.6	18.8	18.9	18.8	19.1	18.8	18.9	18.6	20.2	20.3	19.9	20.6	21.8
600	44.3	45.5	44.7	46.4	44.7	45.0	45.8	48.3	46.1	46.9	46.7	51.8	51.6	50.1	52.6	59.9
700	60.1	62.0	61.6	63.7	61.1	62.7	63.1	66.0	63.6	64.1	63.8	68.6	70.4	69.5	71.2	80.8
800	98.1	102.3	100.0	107.8	100.1	102.2	105.9	112.6	106.7	109.8	109.4	120.1	119.4	117.5	121.3	137.9
900	159.0	167.0	168.0	180.3	164.4	167.1	175.1	187.2	177.2	183.8	185.0	208.1	202.8	200.1	204.2	232.6
1000	173.6	180.4	175.5	188.6	172.5	179.1	185.3	197.1	188.4	191.3	190.9	217.7	212.3	208.4	221.5	245.3
Urban																
100	20.5	20.5	21.2	21.9	20.7	21.1	21.5	22.2	22.0	22.1	22.5	23.4	23.0	23.2	23.5	23.7
200	92.3	92.5	94.3	96.0	93.3	94.1	95.0	96.7	96.3	96.7	97.6	99.9	99.8	100.3	100.6	101.2
300	218.0	217.9	221.4	223.5	220.2	221.3	222.4	224.4	223.7	224.5	226.5	229.9	230.2	230.7	231.1	232.1
400	396.5	395.5	402.5	405.5	401.2	402.9	404.0	406.3	405.7	407.0	409.8	414.9	414.8	415.1	415.6	417.3
500	618.1	618.5	627.0	628.9	621.7	626.3	628.8	629.5	628.8	632.2	639.5	644.2	644.1	644.5	645.0	647.0
600	892.4	892.8	900.5	908.6	895.8	900.3	902.3	911.7	911.0	912.6	913.5	921.9	921.8	924.8	922.7	926.1
700	1230.2	1226.0	1242.3	1244.3	1238.2	1242.2	1243.1	1247.1	1251.2	1256.5	1267.2	1263.5	1268.3	1268.7	1269.3	1274.7
800	1556.5	1556.9	1571.6	1576.6	1570.8	1571.4	1575.3	1577.1	1576.5	1592.6	1607.4	1608.6	1608.6	1609.0	1609.6	1618.3
900	1974.9	1966.3	1976.9	1984.7	1976.1	1976.7	1977.6	1985.1	1984.5	1984.8	1985.7	1999.6	1999.6	2000.0	2000.6	2001.3
1000	2421.8	2422.2	2423.7	2425.7	2422.9	2423.5	2424.4	2426.2	2425.6	2433.3	2453.0	2454.2	2454.2	2454.6	2455.1	2455.9

	Percent of cells with difference of +/- 10 people per hectare from "true", EA-level input data, excluding cells with estimated population <1															
Cell size	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
(metres) **All**																
100	49.1	49.9	52.1	47.4	47.0	44.6	39.8	47.5	44.2	43.9	46.4	44.2	44.2	45.2	42.6	43.2
200	70.7	69.2	71.5	69.4	68.4	67.8	64.5	67.3	67.6	68.2	68.7	67.8	68.5	68.3	69.8	69.7
300	73.5	73.1	73.1	72.3	72.9	72.6	71.8	72.2	73.5	73.5	73.7	73.1	73.9	74.0	73.7	73.4
400	73.7	73.3	73.1	72.6	73.2	73.1	72.9	72.5	73.9	73.8	73.8	73.2	74.3	74.2	74.0	73.6
500	72.9	72.6	72.5	71.7	72.4	72.4	72.1	71.6	72.6	72.7	72.5	72.4	73.8	73.8	73.6	73.4
600	72.7	72.6	72.6	71.9	72.2	72.0	72.0	71.6	72.8	72.6	72.5	71.8	73.3	73.3	72.9	72.6
700	73.2	73.2	73.0	72.5	72.6	72.6	72.5	72.0	73.0	72.9	72.7	72.4	74.0	74.0	74.0	73.8
800	72.9	72.9	72.8	72.1	72.6	72.4	72.4	71.9	73.0	72.8	72.5	72.5	73.5	73.4	73.3	73.0
900	74.2	74.0	73.7	73.1	73.3	73.2	73.1	72.4	74.5	74.3	74.1	73.8	75.6	75.6	75.4	75.2
1000	73.7	73.6	73.0	72.5	72.8	72.7	72.7	72.3	72.4	72.3	72.5	72.4	73.2	73.2	73.2	72.9
Rural																
100	39.6	41.6	45.9	39.1	36.7	32.5	24.3	40.0	32.7	32.3	37.0	34.3	33.2	35.1	31.1	32.8
200	87.8	85.2	90.0	87.5	84.2	83.2	78.2	83.5	81.8	83.1	84.3	83.6	82.7	82.3	85.6	86.0
300	97.3	96.7	97.4	97.0	96.8	96.4	95.6	96.9	96.6	97.0	97.7	97.4	97.3	97.7	97.8	97.8
400	97.9	97.8	97.9	97.8	97.8	97.7	97.6	97.8	98.0	98.0	98.2	98.1	98.3	98.4	98.4	98.3
500	98.2	98.2	98.2	98.2	98.1	98.2	98.2	98.1	98.6	98.6	98.6	98.8	99.1	99.1	99.1	99.3
600	96.4	96.4	96.4	96.3	96.5	96.5	96.5	96.5	96.8	96.8	96.8	96.7	97.5	97.5	97.4	97.4
700	96.2	96.2	96.2	96.1	96.4	96.4	96.3	96.3	96.5	96.5	96.5	96.4	97.7	97.6	97.7	97.6
800	94.9	94.9	94.9	94.8	94.8	94.7	94.7	94.6	95.2	95.2	95.1	95.1	95.9	95.8	95.8	95.8
900	94.1	94.1	94.0	93.8	93.9	93.9	93.8	93.7	95.1	95.0	95.0	94.9	96.4	96.4	96.4	96.4
1000	96.8	96.8	96.6	96.6	96.6	96.6	96.6	96.6	96.4	96.4	96.4	96.4	96.9	96.9	96.9	96.9
Urban																
100	60.0	59.6	59.3	57.1	59.2	58.9	58.2	56.6	59.3	59.3	58.8	57.7	59.5	59.4	58.7	58.1
200	51.2	50.7	49.9	48.2	49.9	49.6	48.8	47.8	49.4	49.1	48.4	46.9	49.4	49.3	48.3	47.2
300	44.7	44.2	43.6	42.1	43.5	43.4	42.6	41.2	42.6	42.2	41.6	40.1	41.6	41.4	40.0	39.0
400	43.0	42.4	42.0	40.8	42.0	42.1	41.6	40.0	41.2	40.8	40.2	38.7	40.3	40.1	39.0	37.7
500	40.6	40.1	39.7	38.3	39.7	39.5	38.8	37.9	38.0	37.9	36.8	36.1	38.7	38.4	36.9	36.2
600	40.5	40.3	40.3	38.5	39.8	39.0	39.0	37.7	38.7	38.3	37.9	36.2	38.0	38.0	36.7	35.9
700	40.2	40.2	40.2	38.8	38.8	38.8	38.8	38.1	38.4	38.1	37.8	37.2	38.5	38.6	38.0	37.6
800	40.6	40.6	40.6	39.0	40.4	40.4	40.0	39.1	40.6	39.5	38.8	38.0	39.1	39.1	37.8	36.4
900	44.5	44.5	44.2	42.7	43.0	43.0	43.0	41.5	43.6	43.2	42.3	41.5	43.5	43.5	42.6	41.3
1000	39.5	39.5	39.1	38.1	38.1	38.1	38.4	36.9	37.1	36.5	36.8	36.3	37.0	37.0	37.0	35.8

Percent of cells with difference of +/- 10 people per hectare from "true" population, Constituency-level input data, excluding cells with estimated population <1																
Cell size (metres)	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
All																
100	61.8	62.4	65.5	70.3	62.8	63.2	66.4	70.3	66.9	68.3	71.8	72.5	69.2	70.9	71.7	71.7
200	76.3	76.0	76.0	75.2	76.1	75.8	75.5	74.5	75.1	74.8	74.8	73.0	73.0	73.2	72.6	70.9
300	74.3	73.9	73.7	72.7	74.0	73.6	73.1	72.1	72.7	72.5	72.4	70.7	70.7	71.1	70.3	68.4
400	72.8	72.3	72.3	71.4	72.4	72.0	71.7	70.8	71.4	71.1	71.0	69.2	69.2	69.5	68.8	67.0
500	71.8	71.3	71.2	70.4	71.5	71.0	70.6	69.8	70.5	70.2	70.1	68.2	68.5	68.6	67.8	65.9
600	70.5	70.0	70.1	69.3	70.2	70.1	69.7	68.6	69.5	69.2	69.2	67.3	67.5	67.8	67.1	64.8
700	70.9	70.2	70.2	69.6	70.5	70.0	69.8	69.2	69.7	69.5	69.5	67.9	67.9	68.1	67.2	65.1
800	70.5	70.1	70.0	69.0	70.3	69.7	69.3	68.3	69.1	68.5	68.3	66.8	66.8	67.1	66.6	64.4
900	71.1	70.4	70.2	69.3	70.5	70.3	69.5	68.7	69.6	69.0	68.9	66.9	67.0	67.2	66.9	64.8
1000	69.7	68.9	69.3	68.2	69.5	69.0	68.3	67.3	68.2	68.2	68.0	65.8	66.4	66.4	65.6	63.8
Rural																
100	71.0	72.6	77.7	87.8	73.2	74.5	80.5	89.0	80.6	83.7	90.7	94.7	88.0	91.0	94.1	96.6
200	99.1	99.1	99.1	99.0	99.2	99.1	99.1	99.0	99.0	99.0	99.0	98.8	98.9	98.9	98.8	98.6
300	98.7	98.6	98.5	98.4	98.5	98.5	98.4	98.3	98.4	98.3	98.3	98.1	98.1	98.2	98.1	97.7
400	98.1	98.1	98.1	98.0	98.1	98.1	98.0	97.9	98.0	97.9	97.9	97.7	97.7	97.8	97.7	97.3
500	98.4	98.3	98.3	98.2	98.3	98.3	98.3	98.1	98.2	98.2	98.2	98.0	98.0	97.8	97.7	97.4
600	96.2	96.0	96.1	95.9	96.1	96.1	95.9	95.7	95.9	95.8	95.8	95.3	95.4	95.5	95.2	94.5
700	95.9	95.4	95.4	95.2	95.4	95.3	95.2	95.0	95.2	95.1	95.2	94.6	94.7	94.7	94.4	93.6
800	94.2	93.7	94.1	93.6	93.9	94.0	93.8	93.3	93.7	93.5	93.5	92.8	92.9	93.0	92.7	91.7
900	92.9	92.6	92.5	92.0	92.7	92.6	92.2	91.7	92.1	91.8	91.8	90.7	91.0	91.1	90.9	89.6
1000	95.9	95.7	95.8	95.5	95.8	95.7	95.5	95.2	95.5	95.4	95.4	94.7	94.9	94.9	94.6	94.0
Urban																
100	52.4	52.3	53.3	53.4	52.2	51.9	52.7	53.1	53.9	53.8	53.7	53.4	53.1	53.0	52.7	53.0
200	52.6	52.9	52.4	52.0	52.4	52.4	52.3	51.9	52.0	51.8	51.6	50.8	50.5	50.3	50.3	50.2
300	48.7	48.9	48.2	47.7	48.2	48.2	48.2	47.8	47.8	47.8	47.3	46.6	46.5	46.5	46.5	46.3
400	46.2	46.2	45.4	45.2	45.5	45.4	45.4	45.3	45.1	44.9	44.7	44.1	44.0	44.0	44.0	43.8
500	44.1	44.1	43.4	43.4	43.8	43.3	43.3	43.4	43.4	43.3	42.8	42.5	42.5	42.5	42.5	42.4
600	42.7	42.7	42.5	42.3	42.5	42.5	42.6	42.1	42.1	42.2	42.2	41.7	41.7	41.5	41.7	41.5
700	43.3	43.6	43.1	43.1	43.3	43.1	43.1	43.2	43.0	42.8	42.4	42.6	42.4	42.4	42.4	42.1
800	44.4	44.4	43.8	44.1	43.8	43.8	44.1	44.1	44.1	43.5	42.9	42.9	42.9	42.9	42.9	42.6
900	45.6	45.9	45.6	45.9	45.6	45.6	45.6	45.9	45.9	45.9	45.9	45.5	44.9	44.9	44.9	44.9
1000	40.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	40.8	40.3	40.3	40.3	40.3	40.3	40.3

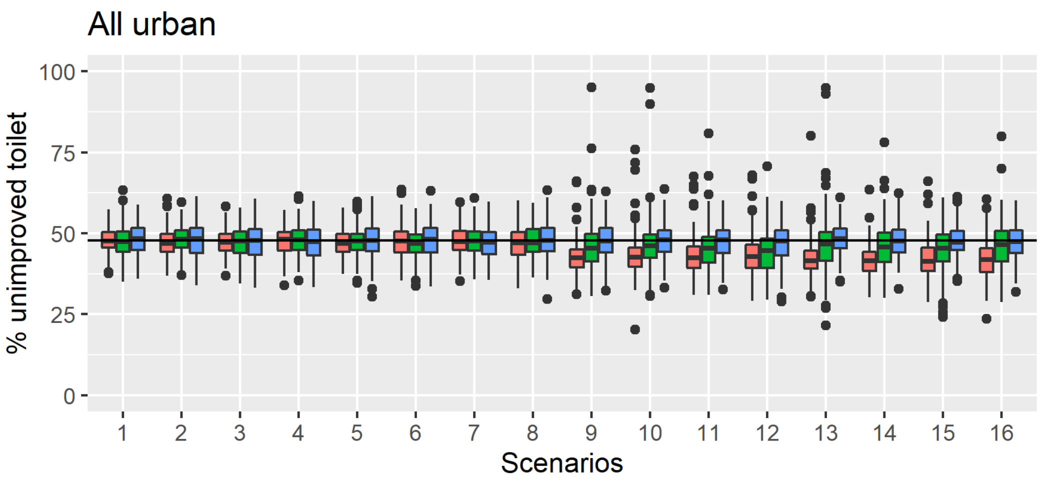
Percent of cells with +/- 20 the "true" population, EA-level input data, excluding cells with estimated population <1																
Cell size	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
(metres) **All**																
100	8.0	8.0	7.3	6.2	7.9	7.8	7.0	6.1	6.9	6.8	6.4	5.7	6.1	5.9	5.6	5.1
200	25.7	24.6	21.5	18.3	24.2	22.8	20.4	17.8	21.5	20.3	18.3	16.5	18.1	17.3	16.0	14.5
300	44.6	44.8	44.6	44.5	43.6	43.6	43.4	43.3	41.9	41.7	41.5	41.1	40.5	40.4	40.1	39.7
400	48.5	48.5	48.2	47.6	47.2	47.1	46.9	46.5	45.4	45.1	44.8	44.3	43.6	43.6	43.1	42.3
500	50.7	51.0	50.5	49.8	49.5	49.6	49.3	48.9	47.7	47.7	47.1	46.5	45.6	45.6	44.8	43.9
600	52.6	52.7	52.3	52.1	51.8	51.5	51.5	50.7	48.8	48.5	48.4	47.7	46.5	46.5	45.6	45.2
700	54.4	54.7	54.4	53.3	52.6	52.2	52.0	51.7	49.8	49.8	50.0	49.3	47.6	47.5	47.1	46.7
800	55.3	55.5	55.4	54.5	54.6	54.3	54.1	53.8	51.9	51.6	51.2	50.7	49.6	49.8	49.0	48.1
900	55.9	55.6	55.4	54.1	54.7	54.6	54.1	53.5	52.1	51.7	51.3	51.2	50.0	50.0	49.5	49.1
1000	56.7	56.6	55.6	55.1	55.0	54.9	54.3	54.0	51.7	51.6	51.8	51.6	49.6	49.6	49.6	49.1
Rural																
100	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1
200	3.8	3.9	3.8	3.1	3.4	3.3	3.2	2.7	2.8	2.8	2.4	2.1	2.2	2.1	1.9	1.7
300	10.8	10.8	10.3	9.7	9.3	9.1	8.7	8.3	7.7	7.4	7.2	6.5	5.8	5.7	5.4	4.6
400	15.2	15.0	14.3	13.2	13.0	12.7	12.3	11.9	10.7	10.2	10.0	8.9	8.5	8.3	7.7	6.5
500	16.9	17.3	16.2	14.5	14.5	14.6	14.1	12.9	11.9	11.9	11.5	10.7	9.4	9.4	9.0	7.2
600	20.7	20.6	20.0	19.6	18.6	18.0	18.2	17.3	15.1	14.7	14.5	12.8	11.6	11.6	10.5	9.6
700	24.6	25.1	24.1	22.6	21.2	20.6	19.9	19.3	17.1	17.1	17.3	15.7	13.4	13.1	13.0	12.2
800	26.2	26.5	26.3	24.3	24.7	23.9	23.6	22.8	19.6	19.7	19.0	19.0	17.0	17.1	16.7	16.0
900	28.5	27.9	27.3	25.1	25.8	25.5	24.3	23.4	21.8	21.0	20.6	20.2	18.6	18.6	18.1	18.1
1000	28.5	28.1	25.7	24.0	25.1	24.7	23.4	23.6	19.3	19.3	19.9	19.9	16.9	16.9	16.9	16.4
Urban																
100	17.0	17.0	15.4	13.2	16.9	16.7	15.1	13.2	15.9	15.5	14.6	13.3	14.4	14.0	13.3	12.1
200	51.8	49.5	42.8	36.8	49.5	46.8	41.6	36.6	46.6	43.9	39.8	36.3	40.5	38.8	36.1	32.8
300	86.6	87.2	87.3	88.1	86.9	87.1	87.3	88.0	88.6	88.5	88.6	89.2	89.5	89.3	89.9	90.3
400	91.5	91.9	91.7	92.0	91.6	91.5	91.7	91.9	93.7	93.8	93.4	94.2	94.3	94.5	95.0	94.9
500	94.2	94.6	94.8	95.1	94.4	94.6	95.0	95.3	96.4	96.3	96.3	96.2	96.4	96.6	96.7	96.9
600	96.8	97.1	96.8	97.3	97.0	97.3	97.0	96.9	97.0	97.0	97.3	98.0	97.9	97.9	97.8	98.2
700	98.0	98.0	98.0	98.0	97.9	97.9	97.9	97.9	98.7	98.7	98.6	99.1	99.5	99.5	99.5	99.5
800	98.4	98.4	98.4	98.9	98.4	98.4	98.9	98.9	99.4	99.4	99.4	99.4	100.0	100.0	100.0	100.0
900	97.4	97.4	97.4	97.3	98.0	98.0	98.0	98.0	98.6	98.6	98.5	99.3	99.2	99.2	99.2	99.2
1000	99.2	99.2	99.2	100.0	99.2	99.2	99.2	99.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

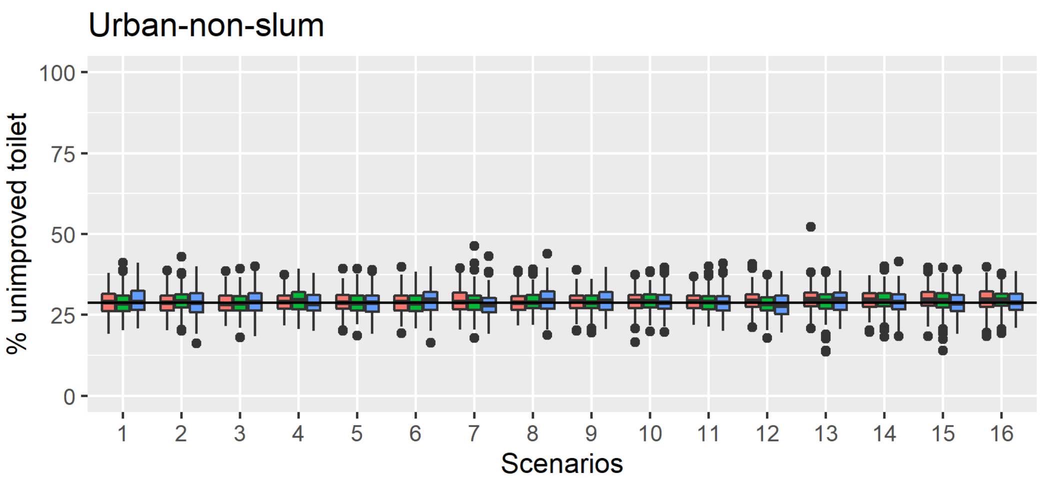
Percent of cells with +/- 20 the "true" population, Constituency-level input data, excluding cells with estimated population <1																
Cell size (metres)	**2016_True**	**2016_L**	**2016_M**	**2016_H**	**2011_True**	**2011_L**	**2011_M**	**2011_H**	**2006_True**	**2006_L**	**2006_M**	**2006_H**	**2001_True**	**2001_L**	**2001_M**	**2001_H**
All																
100	3.4	3.4	3.1	2.8	3.3	3.2	3.1	2.7	2.7	2.6	2.5	2.4	2.3	2.1	2.0	2.1
200	12.5	12.5	11.6	10.8	11.8	11.8	11.6	11.0	10.9	10.7	10.1	10.2	9.5	9.0	9.1	9.1
300	41.4	42.1	42.0	43.3	41.6	42.2	43.0	44.4	43.2	43.7	43.6	46.2	46.1	45.6	46.8	49.2
400	46.1	47.0	46.9	48.1	46.7	47.2	47.8	49.1	48.2	48.4	48.4	51.1	50.9	50.6	51.5	54.3
500	50.5	51.3	51.1	52.5	50.7	51.5	52.1	53.6	52.4	53.0	53.2	56.3	55.7	55.5	56.9	59.5
600	54.3	55.4	55.1	56.8	54.6	55.1	56.0	58.1	56.6	57.2	57.2	60.1	60.0	59.5	60.8	63.7
700	60.2	60.8	60.6	61.5	60.6	61.0	61.4	62.4	61.6	61.9	62.0	64.5	64.7	64.7	65.8	68.8
800	61.7	62.6	62.3	63.8	61.7	62.8	63.7	65.0	63.7	64.8	64.5	66.8	66.8	66.6	67.4	70.9
900	64.0	65.6	65.7	67.4	64.9	65.5	67.1	68.1	66.8	67.4	67.9	70.6	70.3	70.5	70.6	73.7
1000	68.6	70.0	69.0	71.6	68.4	69.7	70.9	72.8	71.6	71.6	71.8	76.1	75.3	74.5	76.8	78.6
Rural																
100	0.2	0.1	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3
200	2.8	2.8	2.8	2.8	2.8	2.8	2.9	3.0	3.0	3.0	2.8	3.2	3.0	2.7	3.1	3.3
300	8.8	9.3	8.9	9.4	8.9	9.1	9.4	9.9	9.4	9.6	9.5	10.6	10.4	10.3	10.8	11.8
400	13.1	13.7	13.7	14.3	13.5	13.6	14.0	14.9	14.4	14.4	14.4	15.4	15.4	15.5	15.7	17.4
500	17.0	17.5	17.4	18.3	17.1	17.5	18.0	18.9	18.0	18.6	18.6	20.8	20.2	20.2	21.3	22.3
600	23.7	24.8	24.6	26.0	23.9	24.4	25.4	27.2	25.9	26.5	26.5	28.8	28.8	28.3	29.5	31.1
700	32.4	32.9	32.7	33.3	32.7	33.0	33.3	34.2	33.3	33.6	33.5	35.5	35.8	36.2	36.9	39.7
800	33.6	35.0	33.9	34.8	33.8	34.3	35.1	35.9	35.0	35.5	34.7	36.5	36.8	36.8	37.4	41.4
900	39.7	41.1	41.4	43.2	40.4	41.1	43.1	43.6	42.4	42.8	43.7	45.7	45.8	46.6	46.2	49.6
1000	43.2	44.6	43.4	46.6	42.4	44.3	45.5	47.6	46.6	46.6	46.2	51.8	50.8	49.2	52.7	54.0
Urban																
100	6.7	6.6	6.1	5.2	6.4	6.2	5.8	5.0	5.1	4.9	4.7	4.4	4.1	3.8	3.5	3.5
200	22.6	22.2	20.6	18.6	21.1	20.8	20.2	18.4	18.5	18.0	17.2	16.3	15.1	14.7	14.1	13.5
300	75.6	75.6	75.9	76.3	75.9	76.0	76.2	76.4	76.2	76.5	76.7	77.4	77.4	77.5	77.5	77.6
400	81.0	80.8	81.4	81.5	81.5	81.5	81.5	81.5	81.5	81.5	81.8	82.4	82.4	82.4	82.4	82.5
500	85.3	85.3	85.7	85.9	85.4	85.9	85.8	85.9	85.9	86.2	86.8	87.1	87.1	87.1	87.1	87.3
600	87.5	87.5	87.7	88.2	87.5	87.7	87.7	88.4	88.4	88.4	88.4	88.9	88.9	89.1	88.9	89.1
700	90.9	90.5	90.8	90.8	90.8	90.8	90.8	90.7	91.1	91.4	92.2	91.8	92.2	92.2	92.2	92.1
800	92.7	92.7	93.1	93.1	93.1	93.1	93.1	93.1	93.1	94.0	94.4	94.4	94.4	94.4	94.4	94.4
900	92.4	92.5	92.4	92.4	92.4	92.4	92.4	92.4	92.4	92.4	92.4	92.9	92.9	92.9	92.9	92.9
1000	96.9	96.9	96.9	96.9	96.9	96.9	96.9	96.9	96.9	96.9	97.7	97.7	97.7	97.7	97.7	97.7

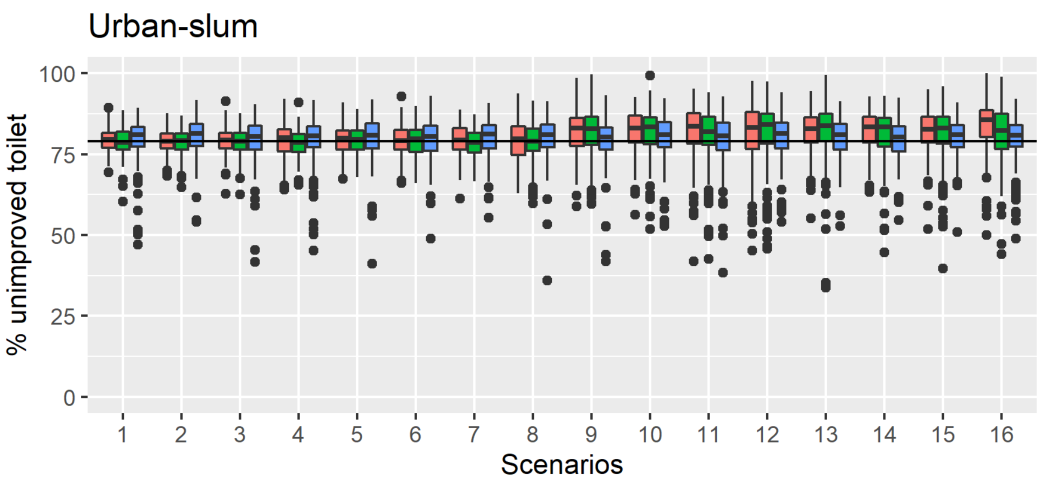
Appendix F Box plots of all survey results from census EA, gridEZ-EA, and gridEZ-constituency sample frames

1	2016_True	**5**	2011_True	**9**	2006_True	**13**	2001_True
2	2016_L	**6**	2011_L	**10**	2006_L	**14**	2001_L
3	2016_M	**7**	2011_M	**11**	2006_M	**15**	2001_M
4	2016_H	**8**	2011_H	**12**	2006_H	**16**	2001_H

Census EA   GridEZ 1 (EA input)   GridEZ 2 (constituency input)   "True" mean

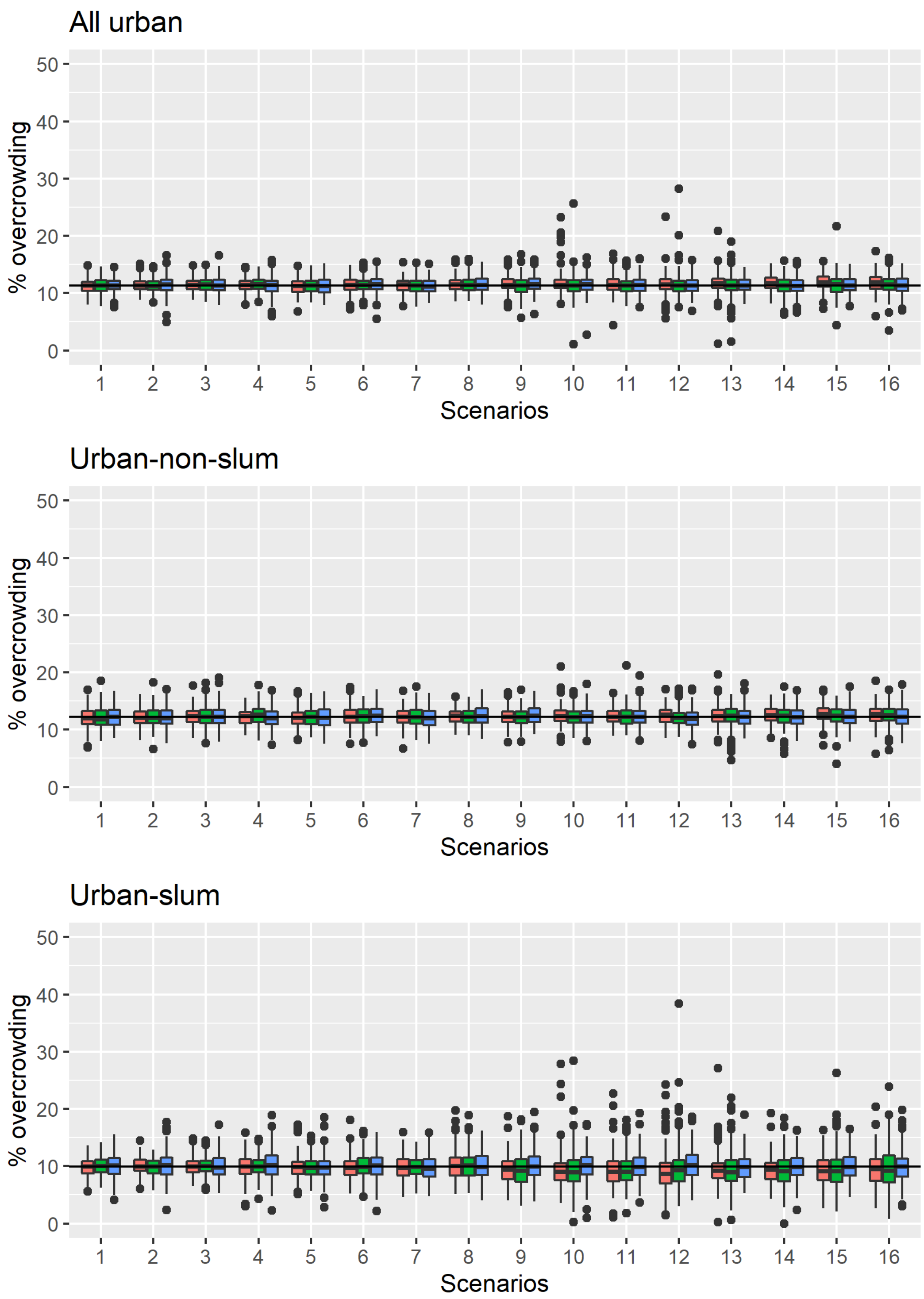




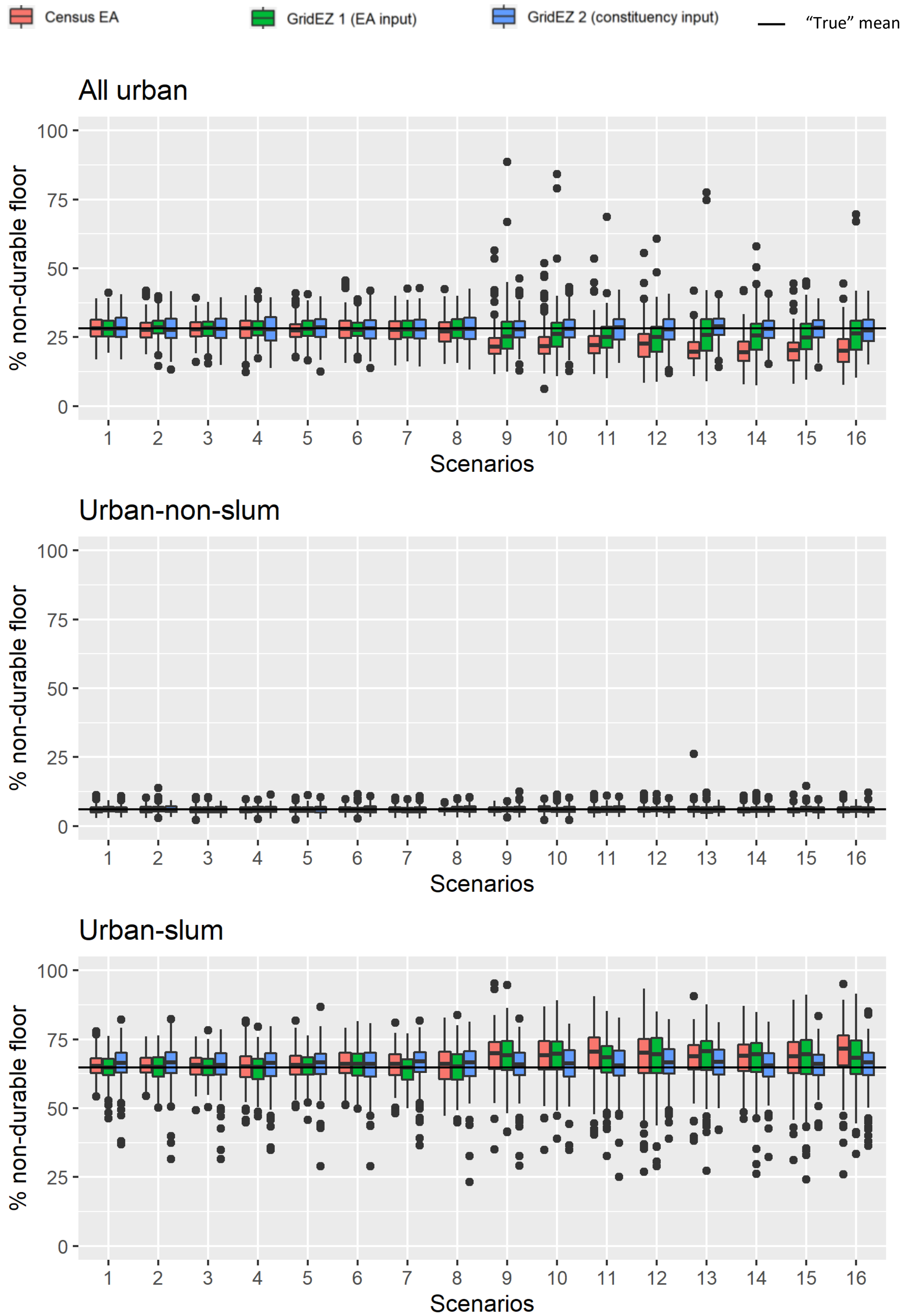


1 2016_True	**5** 2011_True	**9** 2006_True	**13** 2001_True
2 2016_L	**6** 2011_L	**10** 2006_L	**14** 2001_L
3 2016_M	**7** 2011_M	**11** 2006_M	**15** 2001_M
4 2016_H	**8** 2011_H	**12** 2006_H	**16** 2001_H

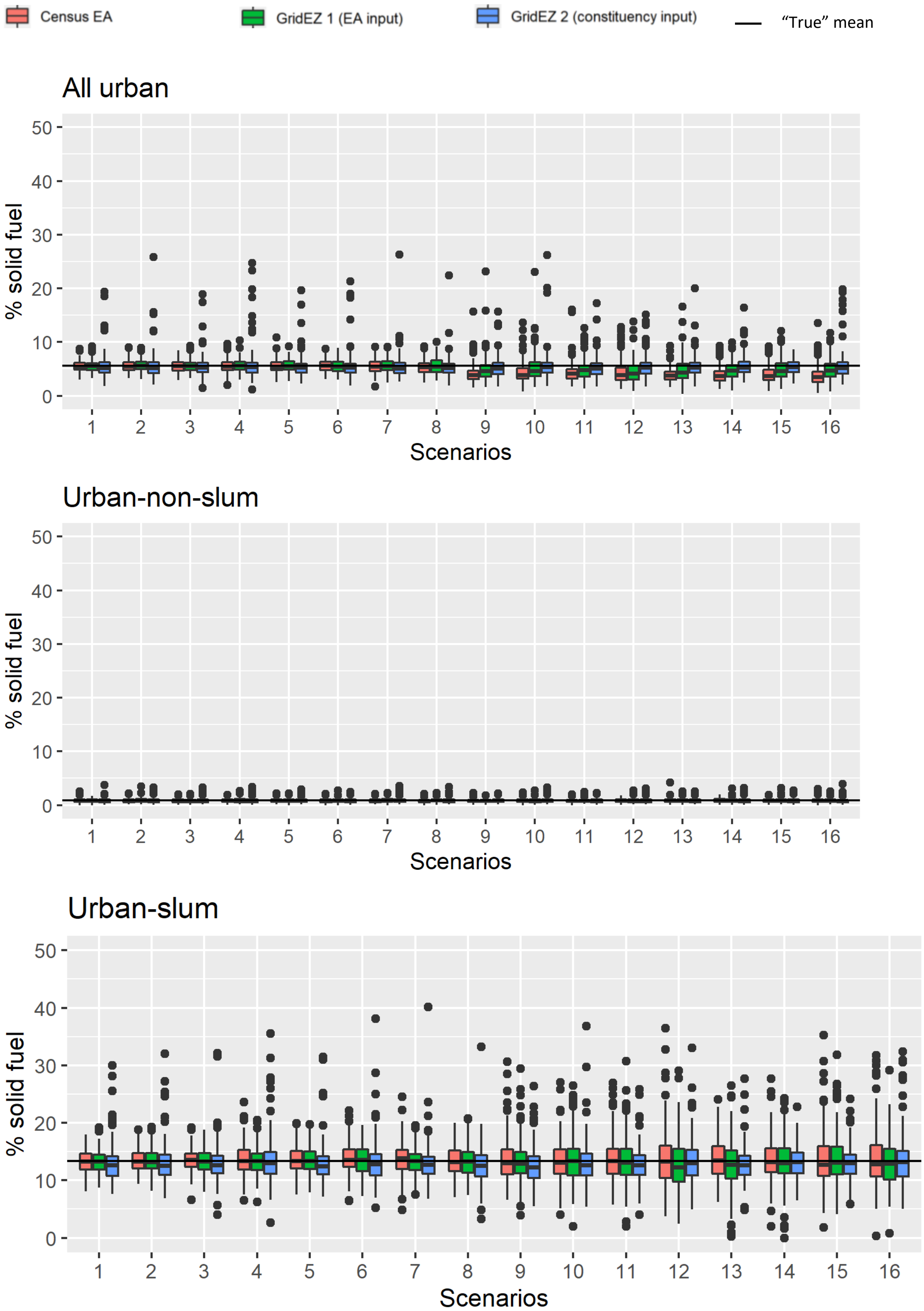
Census EA | GridEZ 1 (EA input) | GridEZ 2 (constituency input) | "True" mean



1 2016_True
2 2016_L
3 2016_M
4 2016_H
5 2011_True
6 2011_L
7 2011_M
8 2011_H
9 2006_True
10 2006_L
11 2006_M
12 2006_H
13 2001_True
14 2001_L
15 2001_M
16 2001_H



1	2016_True	**5**	2011_True	**9**	2006_True	**13**	2001_True
2	2016_L	**6**	2011_L	**10**	2006_L	**14**	2001_L
3	2016_M	**7**	2011_M	**11**	2006_M	**15**	2001_M
4	2016_H	**8**	2011_H	**12**	2006_H	**16**	2001_H



Appendix G RMSE and bias statistics for all survey results

RMSE	Census			gridEZ-EA			gridEZ-Constituency		
	All	Non-slum	Slum	All	Non-slum	Slum	All	Non-Slum	Slum
Scenario				% unimproved toilet					
2016_True	3.9	3.8	3.4	4.4	3.6	4.2	4.9	3.8	3.9
2016_L	4.1	3.5	3.5	4.2	3.6	4.1	5.3	4.3	4.1
2016_M	4.0	3.1	3.8	4.4	3.3	4.2	5.3	4.0	4.0
2016_H	4.3	3.1	5.1	4.6	3.7	4.3	5.2	3.6	4.3
2011_True	4.0	3.4	4.4	4.2	3.4	4.2	5.3	3.9	4.0
2011_L	4.6	3.3	4.8	4.2	3.4	4.7	5.2	3.9	4.6
2011_M	4.7	3.4	4.8	4.5	3.7	4.4	5.0	3.9	4.7
2011_H	4.8	3.1	6.0	4.7	3.1	5.6	5.3	4.2	4.8
2006_True	7.0	3.2	6.8	7.6	3.0	7.6	5.3	3.9	7.0
2006_L	7.6	3.1	7.2	7.7	3.3	8.0	5.2	3.4	7.6
2006_M	7.2	3.1	8.8	6.7	3.3	8.4	5.3	3.6	7.2
2006_H	7.4	3.5	9.7	7.4	3.1	9.4	5.9	4.1	7.4
2001_True	8.1	3.9	6.6	9.0	3.9	9.1	4.8	3.8	8.1
2001_L	7.8	3.2	7.1	6.9	3.6	7.9	4.9	3.7	7.8
2001_M	8.0	3.5	7.1	7.1	3.8	8.6	5.0	3.6	8.0
2001_H	8.3	3.7	9.0	7.3	3.3	9.0	5.3	3.3	8.3
Scenario				% overcrowding					
2016_True	1.3	1.8	1.5	1.2	1.7	1.6	1.3	1.8	1.3
2016_L	1.1	1.5	1.5	1.1	1.6	1.4	1.5	1.8	1.1
2016_M	1.1	1.5	1.6	1.3	1.8	1.6	1.4	1.9	1.1
2016_H	1.1	1.3	2.1	1.2	1.8	1.7	1.6	1.7	1.1
2011_True	1.3	1.5	1.9	1.2	1.6	1.8	1.5	1.9	1.3
2011_L	1.3	1.6	1.8	1.3	1.6	2.1	1.4	1.7	1.3
2011_M	1.2	1.6	2.1	1.3	1.7	1.7	1.3	1.7	1.2
2011_H	1.3	1.3	2.5	1.4	1.5	2.4	1.5	1.8	1.3
2006_True	1.4	1.5	2.3	1.6	1.5	2.9	1.5	1.7	1.4
2006_L	1.8	1.5	3.0	1.9	1.5	3.2	1.4	1.7	1.8
2006_M	1.5	1.4	2.9	1.5	1.7	2.6	1.4	1.7	1.5
2006_H	1.7	1.5	3.4	2.0	1.4	3.9	1.4	1.7	1.7
2001_True	1.7	1.8	2.8	2.0	1.8	3.3	1.3	1.7	1.7
2001_L	1.4	1.5	2.4	1.5	1.7	3.0	1.5	1.7	1.4
2001_M	1.6	1.6	2.7	1.7	1.7	3.1	1.3	1.7	1.6
2001_H	1.6	1.6	3.0	1.6	1.6	3.6	1.5	1.8	1.6
Scenario				% non-durable floor					
2016_True	4.0	1.4	4.3	4.0	1.2	4.9	4.9	1.4	4.0
2016_L	4.1	1.4	4.1	4.1	1.5	5.0	5.2	1.3	4.1
2016_M	4.0	1.3	4.5	4.1	1.3	4.5	4.9	1.3	4.0
2016_H	4.8	1.2	6.3	4.3	1.4	5.2	5.3	1.3	4.8
2011_True	4.2	1.4	5.4	4.1	1.2	4.8	4.9	1.4	4.2
2011_L	4.6	1.2	5.8	4.0	1.3	6.0	4.9	1.5	4.6
2011_M	5.0	1.3	6.3	4.3	1.2	5.5	5.0	1.3	5.0

2011_H	5.0	1.1	7.2	4.9	1.2	7.1	5.4	1.5	5.0
2006_True	8.2	1.2	8.6	8.3	1.1	9.0	5.1	1.5	8.2
2006_L	8.1	1.3	8.9	8.7	1.3	9.0	5.5	1.3	8.1
2006_M	7.9	1.3	10.0	6.9	1.2	8.8	5.3	1.4	7.9
2006_H	8.7	1.4	11.2	8.2	1.2	10.7	5.8	1.4	8.7
2001_True	9.2	2.0	8.1	8.8	1.5	10.0	4.8	1.4	9.2
2001_L	9.5	1.3	8.4	7.7	1.3	9.6	4.8	1.3	9.5
2001_M	9.7	1.3	9.1	7.2	1.3	10.6	4.8	1.4	9.7
2001_H	10.1	1.5	11.1	8.3	1.3	10.4	5.3	1.4	10.1
Scenario					% solid fuel				
2016_True	1.1	0.4	1.9	1.1	0.4	1.9	2.1	0.4	1.1
2016_L	1.1	0.4	2.0	1.1	0.4	2.1	2.3	0.5	1.1
2016_M	1.1	0.4	2.0	1.1	0.4	2.0	1.9	0.4	1.1
2016_H	1.2	0.3	2.8	1.2	0.4	2.2	2.9	0.5	1.2
2011_True	1.1	0.4	2.4	1.2	0.4	2.1	1.9	0.4	1.1
2011_L	1.2	0.4	2.4	1.2	0.4	2.6	2.3	0.5	1.2
2011_M	1.3	0.4	2.7	1.2	0.4	2.2	2.1	0.5	1.3
2011_H	1.4	0.4	2.7	1.4	0.3	2.9	1.9	0.5	1.4
2006_True	2.2	0.4	3.6	2.4	0.4	3.8	1.8	0.4	2.2
2006_L	2.2	0.4	3.4	2.7	0.4	4.0	2.5	0.5	2.2
2006_M	2.2	0.4	3.5	2.0	0.3	4.0	1.8	0.4	2.2
2006_H	2.4	0.4	4.9	2.3	0.4	4.4	1.9	0.4	2.4
2001_True	2.1	0.4	3.5	2.4	0.4	4.1	1.8	0.4	2.1
2001_L	2.2	0.4	3.8	1.9	0.4	3.9	1.7	0.5	2.2
2001_M	2.4	0.3	4.6	2.0	0.4	4.4	1.3	0.5	2.4
2001_H	2.6	0.4	5.1	2.1	0.4	4.2	2.6	0.5	2.6

Bias	Census			gridEZ-EA			gridEZ-Constituency		
	All	Non-slum	Slum	All	Non-slum	Slum	All	Non-Slum	Slum
Scenario	% unimproved toilet								
2016_True	0.0	-0.1	0.4	0.0	0.0	-0.1	0.4	0.4	0.9
2016_L	-0.4	0.0	0.1	0.5	0.6	-0.2	0.2	0.2	1.8
2016_M	-0.4	-0.2	0.2	-0.3	-0.1	-0.1	0.0	0.3	0.6
2016_H	-0.1	0.1	0.5	0.3	0.6	-0.5	-0.3	-0.2	0.7
2011_True	-0.5	0.0	0.5	-0.1	-0.1	0.4	0.1	-0.1	1.2
2011_L	-0.2	0.1	0.3	-0.7	-0.2	0.3	0.1	0.5	0.8
2011_M	-0.1	0.4	0.3	-0.2	0.2	-0.6	-0.6	-0.6	1.3
2011_H	-0.7	0.0	0.3	-0.1	0.1	0.1	0.3	0.8	1.1
2006_True	-5.0	0.3	2.7	-1.9	-0.1	2.5	0.0	0.7	0.4
2006_L	-4.5	0.4	3.3	-1.4	0.4	2.6	0.1	0.2	1.4
2006_M	-4.8	0.2	3.0	-2.3	0.2	2.2	-0.5	0.1	0.5
2006_H	-4.5	0.7	2.4	-3.5	-0.4	2.7	-0.8	-0.6	1.5
2001_True	-5.8	1.0	3.2	-1.5	0.2	3.4	0.6	0.7	1.4
2001_L	-6.1	0.7	3.3	-1.6	0.8	2.5	0.0	0.2	0.8
2001_M	-5.8	0.9	3.1	-2.3	0.4	2.4	-0.1	-0.1	1.3
2001_H	-5.9	0.9	4.9	-1.3	0.7	2.1	-0.4	0.1	0.7
Scenario	% overcrowding								
2016_True	0.0	-0.1	-0.1	0.0	-0.1	0.2	0.1	0.1	0.1
2016_L	0.1	0.0	0.1	0.0	0.0	-0.1	0.1	0.0	0.2
2016_M	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.0
2016_H	0.1	0.0	0.0	0.3	0.4	0.0	0.0	-0.1	0.4
2011_True	-0.1	-0.2	-0.1	0.0	0.0	-0.1	-0.1	-0.1	-0.2
2011_L	0.1	0.1	-0.2	0.2	0.2	0.0	0.3	0.4	0.0
2011_M	0.0	0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.2	-0.2
2011_H	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.2
2006_True	0.3	0.1	-0.5	0.0	-0.1	-0.5	0.4	0.3	0.3
2006_L	0.4	0.2	-0.6	0.0	0.0	-0.6	0.2	0.2	0.2
2006_M	0.2	0.1	-0.7	0.0	0.1	-0.6	0.1	0.1	0.0
2006_H	0.2	0.2	-1.0	0.2	0.0	-0.3	0.1	-0.2	0.4
2001_True	0.4	0.3	-0.6	0.0	0.1	-0.5	0.1	0.1	-0.1
2001_L	0.5	0.4	-0.7	0.0	0.1	-0.7	0.0	0.0	0.0
2001_M	0.7	0.6	-0.7	0.2	0.2	-0.5	0.1	0.2	0.0
2001_H	0.5	0.4	-0.6	0.2	0.3	-0.4	0.1	0.1	0.0
Scenario	% non-durable floor								
2016_True	0.0	-0.2	0.5	0.0	0.1	-0.2	0.3	0.1	1.0
2016_L	-0.4	0.0	0.5	0.3	0.2	-0.2	-0.1	0.0	1.3
2016_M	-0.3	-0.1	0.4	-0.2	-0.1	0.1	-0.2	0.0	0.7
2016_H	-0.5	-0.1	0.3	0.0	0.1	-0.4	-0.2	0.0	0.5
2011_True	-0.6	0.0	0.8	-0.1	0.0	0.4	0.1	0.1	0.9
2011_L	0.0	-0.1	1.2	-0.4	0.0	0.8	-0.3	0.0	0.5
2011_M	-0.4	0.0	0.8	-0.5	-0.2	-0.7	-0.3	-0.1	1.3
2011_H	-0.9	-0.1	0.4	-0.3	0.1	-0.2	-0.2	0.1	1.1
2006_True	-5.9	0.0	4.1	-2.0	-0.1	3.6	-0.2	0.1	0.7
2006_L	-5.5	0.1	4.1	-1.7	0.0	3.8	-0.1	0.1	1.3
2006_M	-5.6	0.0	4.4	-2.8	0.1	2.6	-0.5	0.1	0.5

2006_H	-5.6	0.0	3.4	-3.7	0.0	3.3	-0.6	-0.1	1.4
2001_True	-7.8	0.1	3.1	-2.0	-0.1	3.9	0.4	0.1	1.5
2001_L	-8.1	-0.2	3.2	-2.4	0.0	3.0	-0.2	0.1	0.4
2001_M	-7.9	-0.3	3.4	-2.9	0.1	3.1	-0.2	0.0	1.0
2001_H	-7.9	-0.1	5.3	-1.7	0.1	3.0	-0.7	-0.1	0.5
Scenario				% solid fuel					
2016_True	-0.1	0.0	-0.1	-0.1	0.0	-0.1	-0.1	0.0	-0.5
2016_L	-0.1	0.0	0.1	0.0	0.0	0.0	-0.2	0.0	-0.5
2016_M	-0.1	0.0	0.2	-0.1	-0.1	0.0	-0.2	0.0	-0.3
2016_H	-0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0
2011_True	-0.1	0.0	0.2	0.0	0.0	0.2	-0.2	0.0	-0.5
2011_L	0.1	0.0	0.5	-0.1	0.0	0.1	-0.1	0.0	-0.4
2011_M	-0.1	0.0	0.2	0.0	0.0	0.0	-0.3	0.0	-0.5
2011_H	-0.3	-0.1	0.0	-0.1	0.0	-0.1	-0.4	0.0	-0.7
2006_True	-1.5	0.0	0.0	-0.7	0.0	-0.2	-0.4	0.0	-0.9
2006_L	-1.4	0.0	-0.1	-0.4	0.0	0.2	-0.1	0.0	-0.5
2006_M	-1.3	0.0	0.3	-0.8	0.0	0.0	-0.4	0.0	-0.7
2006_H	-1.3	0.0	0.1	-1.2	0.0	-0.7	-0.3	0.0	-0.1
2001_True	-1.7	0.0	0.4	-0.8	0.0	-0.5	-0.2	0.0	-0.6
2001_L	-1.7	0.0	0.2	-0.8	0.0	-0.1	-0.1	0.0	-0.2
2001_M	-1.7	0.0	0.3	-0.6	0.0	0.4	-0.3	0.0	-0.4
2001_H	-1.9	0.0	0.2	-0.7	0.0	-0.3	-0.1	0.0	0.0

Appendix H Sample frame scale exercise used to decide "seed" cell size in SUE household surveys

1. Use the GridSample R package following guidance in Thomson *et al.*, 2017 and the below code to generate two samples based on 100m and 200m grid cells respectively using the following parameters. We provide this example for Dhaka.

```
library(gridsample)
library(raster)
library(rgdal)
setwd("C:/User/Project")
# Unzip and save downloaded files to subfolder called "data"

clip<-function(raster,shape) {
  a1_crop<-crop(raster,shape)
  step1<-rasterize(shape,a1_crop)
  a1_crop*step1}


#### STRATA SHP ####  --  https://gadm.org/download_country_v3.html
strata <- readOGR(dsn='data', layer='NPL_adm4', encoding='ESRI Shapefile')
proj4string(strata)
strata <- strata[strata@data$NAME_2=="Bagmati",]
plot(strata)


#### POPULATION RASTER ####
#--From https://www.worldpop.org/project/categories?id=3
population_raster <- raster("data/npl_ppp_2020.tif")
proj4string(population_raster)
population_raster <- clip(population_raster,strata)
plot(population_raster)


#### STRATA RASTER ####
strata_raster <- rasterize(strata,population_raster, field="ID_4")
plot(strata_raster)


#### URBAN RASTER ####
#--https://ghsl.jrc.ec.europa.eu/ghs_smod.php
urban_raster <- raster("data/GHS_SMOD_POP2015_GLOBE_R2016A_54009_1k_v1_0.tif")
proj4string(urban_raster)
urban_raster <- projectRaster(urban_raster,crs="+proj=longlat +datum=WGS84
+no_defs +ellps=WGS84 +towgs84=0,0,0")
urban_raster <- clip(urban_raster,strata)
plot(urban_raster)
#Alternatively create a dummy dataset (because reprojection takes ages)
urban_raster <- strata_raster
urban_raster[urban_raster==10] <- 1
plot(urban_raster)


#### DRAW SAMPLE (from 100m cells) ####
psu_polygons=gs_sample(
  population_raster = population_raster,
  strata_raster = strata_raster,
  urban_raster = urban_raster,
  cfg_random_number = 2017,
  cfg_desired_cell_size = 1, # 100m grid cells
  cfg_hh_per_stratum = 200, # 200/20 = 10 PSUs
  cfg_hh_per_urban = 20,
  cfg_hh_per_rural = 20,
  cfg_min_pop_per_cell = 0,
  cfg_max_psu_size = NA,
  cfg_pop_per_psu = 82, # Assumes 20 HHs & ave HH size = 4.1
  cfg_psu_growth = TRUE,
  cfg_sample_rururb = TRUE,
  cfg_sample_spatial = FALSE,
  cfg_sample_spatial_scale = ,
```

```
  output_path=" C:/User/Project",
  sample_name="NPL_100m_sample")
plot(psu_polygons)
#### DRAW SAMPLE (from 200m cells) ####
psu_polygons=gs_sample(
  population_raster = population_raster,
  strata_raster = strata_raster,
  urban_raster = urban_raster,
  cfg_random_number = 2017,
  cfg_desired_cell_size = 2, # 200m grid cells
  cfg_hh_per_stratum = 200, # 200/20 = 10 PSUs
  cfg_hh_per_urban = 20,
  cfg_hh_per_rural = 20,
  cfg_min_pop_per_cell = 0,
  cfg_max_psu_size = NA,
  cfg_pop_per_psu = 82, # Assumes 20 HHs & ave HH size = 4.1
  cfg_psu_growth = TRUE,
  cfg_sample_rururb = TRUE,
  cfg_sample_spatial = FALSE,
  cfg_sample_spatial_scale = ,

output_path="C:/Users/drt1g15/Dropbox/Work_main/p_global_gridsample_feasibility/
bangladesh/thesis_appendix_exercise",
  sample_name="NPL_200m_sample")
plot(psu_polygons)
```

2. Complete the "Sample frame scale worksheet" below with the output shapefile, following the 5 instructional steps below.

   Remember, we are still at the beginning of the sample workflow. This is what we know:
   - Some buildings are not residential
   - Buildings might have multiple levels
   - There may be multiple dwellings on each level
   - There may be multiple households in each dwelling

   This is what we do not know:
   - A complete picture of building locations (in OpenStreetMap)
   - How many dwellings or households are in each PSU

   So, we are going to make an educated guess based on Google Earth imagery, including apparent height of buildings.

   Step 1: Copy the WorldPop population estimate for each PSU into the form.

   After submitting your GridSample job, the user receives an email with a link to download the sample PSU boundaries. The download includes a shapefile (comprised of multiple subfiles including a dbf) which can be opened in ArcGIS or excel, and a kml file which can be opened in Google Earth.

   Download and unzip this file, then open the dbf in Excel. Copy the PSU population estimates (variable name: psu_pop) into the worksheet, and round population to whole numbers.

   Step 2: Use the average household size (entered in How To: Use GridSample exercise) to estimate the number of households per PSU. Note:

   PSU households = PSU population ÷ average household size

Step 3: Open the kml files in Google Earth, and estimate the number of single family and multi-family buildings.

You may need to download Google Earth, first, from: www.google.com/earth/download/gep/agree.html

Once Google Earth is installed on your computer, you only need to double click on the kml file to open it in Google Earth.

Step 4: Look closely at the imagery for each PSU, and answer the following question by circling YES or NO. Does Google Earth appear to show a similar # of HHs as WorldPop? You can answer NO if the imagery appears to have far more or far fewer households than WorldPop.

Step 5. Looking at Google Earth imagery, answer the following question by circling YES or NO: Do you think there are 20+ households in this PSU?

3. This guidance is subject to change as we learn more about gridded population sampling, but here are our initial recommendations.

   - Feasibility/accuracy: If you circled 9-10 YESs in step 4, then this sample frame scale is appropriate for area-microcensus sampling
   - Feasibility: If you expect to find more than 100 households in 4+ PSUs of the 200m X 200m sample frame, then use a 100m X 100m sample frame
   - Accuracy: If you circled NO in step 4 in 4+ PSUs of the 100m X 100m sample frame, then use a 200m X 200m sample frame. (Although this might result in far more households than desired, it may be preferable to aggregate the sample frame to 200m X 200m cells to improve accuracy and then manually segment PSUs, than to sample from an inaccurate sample frame.)

PSU #	**Steps 1 and 2.** WorldPop	**Step 3:** Google Earth and OpenStreetMap	**Step 4.** Does Google Earth show a similar # of HHs as WorldPop?	**Step 5.** Do you think there are 20+ households in this PSU?
First sample with 100m X 100m grid cells				
1	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
2	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
3	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
4	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
5	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
6	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
7	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
8	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
9	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
10	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
Second sample with 200m X 200m grid cells				
1	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
2	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
3	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
4	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
5	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
6	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
7	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
8	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
9	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO
10	Population:______ Households:_____	# singlefam blgs:___ # multifam blgs:____	YES / NO	YES / NO

Appendix I SUE mapper-lister focus group discussion guide

	SUE Survey Focus Group Discussion Guide
FGD Session	
Location	
Date (MM/DD/YY)	
Facilitator's name	
Note taker's name	
Start time	
End time	

Participant ID	***FGD***	***Name***	***Signature***
Participant #1	A		
Participant #2	A		
Participant #3	A		
Participant #4	A		
Participant #5	A		
Participant #6	A		
Participant #7	B		
Participant #8	B		
Participant #9	B		
Participant #10	B		
Participant #11	B		
Participant #12	B		
Participant #13	B		

FGD guide

Core Questions	Probe Questions
Theme A: Introduction	
I would like to start with your overall impressions.	
1. How was your experience on the SUE survey?	
Theme B: Office Enumeration	
I would like to discuss about the pre-field work.	
2. We started with office enumeration in OpenStreetMap. How did OpenStreetMap data help or hinder your work?	• How did OSM data compare across tools (Paper map, GeoODK, OSMAnd)?
Theme C: Field Work	
Now I would like to discuss about your field work experience.	
3. The training provided an idealized script which might have been adapted in practice. What was the "script" that you and your partner used with residents and neighbors?	• How did you approach people? • How did people respond?
4. Reflect on interacting with the residents and neighbors – how did your approach change over time and why?	• What are some tips for talking to different kinds of residents?
5. How would you describe your team's process of enumerating and listing?	• Did you need to adapt how you used OSMAnd, GeoODK, CamScanner, and MAPS.ME compared to the training? • How did you adapt the methods presented in the training in practice? • What roles did you and your partner settle on?
6. Please tell us about the major differences between working in area-microcensus versus two-stage sampling areas.	• Did you enjoy working in one type of sampling area more? • Was there anything particularly challenging about working in one/two-stage sampling areas?
7. Tell us about any field experiences that you had which you did not anticipate at the start.	• What was that experience like for you? • Was this a regular experience? • In your opinion, should anything be incorporated into future trainings to prepare enumerator-listers for this experience?
8. How useful were the regular office meetings and field visits by the Planning Team?	• Did you change your processes after office meetings or field visits? What changed?

9. In other surveys, teams commonly commute to the field from the office or guesthouse, and you commuted directly from your home. Please tell us about your commutes to/from the sampling areas.	• What modes of transport did you use? • How long did it take? • Would you have preferred a different commuting process? If so – what process would you recommend?

Theme D: Workflow & General Reflections

Now I would like to discuss the overall workflow and your general reflections.

10. Think back to the training – what parts best prepared you for the fieldwork, and what was missing from the training?	• Is there anything you would change about the toolkit? • What would you have liked to spend more time on during the training?
11. In typical household surveys, buildings are enumerated using hand-drawn paper maps, and households are listing on paper forms. What do you think were the advantages and disadvantages of the SUE survey methods compared to these typical methods?	• What would it have been like to draw enumeration maps by hand in your PSUs? And to list households on paper forms only?
12. Please tell us about what is was like to work with your field partner.	• Did you have different strengths? Different approaches? • How did any differences play out for you? • Did you and your partner work differently together over time? • Do you have suggestions to improve team work?
13. What do you think the impacts are outside of this project of updating OSM?	
14. As you know we followed three steps – office enumeration, field listing, post-field update of OSM. Is there anything you would change about the workflow?	
15. Would you like to add something to our discussion?	• What additional support might be needed?

Thank you very much for your time!

Appendix J SUE mapper-lister focus group discussion coding framework

Training and Hiring

- Candidate background
- Curriculum content
- Teaching methods
- Contracts
- Misunderstanding

Office work

Interactions

- Partner
- Residents
    - Things neighbours said
    - Things residents said
    - Script
- HERD staff

Field effort

- Transportation or accommodation
- Time of day
- Weather
- One stage vs two stage
- Recording data
- Road or building identification
- Surprises

Field-support

- Visits from planning team
- Office meetings
- Calls to planning team
- Equipment (map, tablet) or supplies [during analysis, this was split: “Paper Maps”, “Tablets”)
- Software on tablet
    - CamScanner
    - Google Maps
    - OSMAnd
    - GeoODK
    - Maps.me
- Satellite imagery sources
    - ArcGIS imagery
    - OSM
    - Paid high-res imagery

General experience

- Recommendation
    - Tip for mappers-listers
- Challenge
- Overall workflow

Glossary of Terms

ancillary variable - A variable that is possibly predictive of the outcome under study. Also called a covariate.

bagging - A statistical technique of drawing multiple random samples with replacement from a dataset during model building to estimate the prediction error of that model. Also called bootstrapping.

big data - Extremely large datasets that are often analysed with machine learning or other computational analysis methods. Includes mobile phone call detail records (CDRs), twitter messages, and satellite imagery.

bootstrapping - A statistical technique of drawing multiple random samples with replacement from a dataset during model building to estimate the prediction error of that model. Also called bagging.

dasymetric - A spatial disaggregation technique that generates a weights layer in smaller sub-areas which informs an unequal disaggregation. The weights layer is generated from ancillary data related to unequal distribution of the original aggregated total. For example, land cover type is often used to dasymetrically disaggregate census population counts to smaller sub-areas.

ensemble - A modelling approach of running two or more related but different models, and synthesizing results into a single value to improve accuracy. Commonly used in "big data" analysis.

geo-statistical - A class of statistical models used to analyse and predict values associated with a spatial or spatiotemporal phenomena.

inverse proportional weighting (IPW) - A modelling technique to calculate statistics standardized to a population different from that which the data were collected. For example, standardizing population characteristics measured in a household survey sample to population characteristics measured in a census.

Markov Chain Monte Carlo (MCMC) - A simulation modelling technique to estimate an outcome by drawing many samples from a large dataset, where each sample informs the next (a Markov chain).

machine learning - Describes a type of computational algorithm which "learns" patterns in very large datasets and use this "acquired" information to improve predictions from the data.

pycnophalactic - A disaggregation technique that preserves the original aggregated total.

Random Forest - An ensemble, machine learning modelling technique that first constructs a multitude of classification "trees" (or models) from a random selection of data, then second runs regressions on each of the trees with the remaining data to estimate a multitude of means and errors, and finally predicts an outcome and model error by averaging.

raster - Refers to coverage datasets whereby data are organized as rectangular, parallel cells (or pixels), and each cell has a single value. Examples include a photograph where each cell has a colour value, temperature data where each cell has an average temperature value for a place and time period, and land cover data where each cell is classified by the dominant land cover type.

List of References

Agarwal, S. (2011) 'The state of urban health in India: Comparing the poorest quartile to the rest of the urban population in selected states and cities', *Environment and Urbanization*, 23(1), pp. 13–28. doi: 10.1177/0956247811398589.

Agarwal, S. and Taneja, S. (2005) 'All slums are not equal: Child health conditions among the urban poor', *Indian pediatrics*, 42(3), pp. 233–244. doi: 10.17763/haer.41.1.91367v0h80051573.

Ahonsi, B. A. (1988) 'Deliberate falsification and census-data in Nigeria', *African Affairs*, 87(349), pp. 553–562.

Alba, R., Muller, W. and Schimpl-Neimanns, B. (1994) 'Secondary analysis of official microdata', in Borg, I. and Mohler, P. P. (eds) *Trends and perspectives in empirical social research*. Berlin Germany: Walter de Gruyter, pp. 57–78.

Alegana, V. A. *et al.* (2016) 'Advances in mapping malaria for elimination: Fine resolution modelling of Plasmodium falciparum incidence', *Scientific Reports*. Nature Publishing Group, 6, p. 29628. doi: 10.1038/srep32908.

Alfons, A. *et al.* (2011) 'Simulation of close-to-reality population data for household surveys with application to EU-SILC', *Statistical Methods and Applications*, 20(3), pp. 383–407. doi: 10.1007/s10260-011-0163-2.

Alkire, S. and Samman, E. (2014) *Mobilizing the household data required to progress toward the SDGs*. Oxford. Available at: http://unsdsn.org/wp-content/uploads/2014/09/Mobilising-the-household-data-required-to-progress-toward-the-SDGs-WEB.pdf.

American Red Cross (2017) *OpenMapKit*. Available at: http://openmapkit.org/ (Accessed: 13 January 2020).

Anderssen, N. and Malterud, K. (2017) 'Oversampling as a methodological strategy for the study of self-reported health among lesbian, gay and bisexual populations', *Scandinavian Journal of Public Health*, 45(6), pp. 637–646. doi: 10.1177/1403494817717407.

Angeles, G. *et al.* (2009) 'The 2005 census and mapping of slums in Bangladesh: Design, select results and application', 8, pp. 1–19. doi: 10.1186/1476-072X-8-32.

Azar, D. *et al.* (2010) 'Spatial refinement of census population distribution using remotely sensed estimates of impervious surfaces in Haiti', *International Journal of Remote Sensing*, 31(21), pp. 5635–5655. doi: 10.1080/01431161.2010.496799.

Azar, D. *et al.* (2013) 'Generation of fine-scale population layers using multi-resolution satellite imagery and geospatial data', *Remote Sensing of Environment*, 130, pp. 219–232. doi: 10.1016/j.rse.2012.11.022.

Balk, D. *et al.* (2005) *Mapping global urban and rural population distributions*. 24. Rome Italy. Available at: http://www.fao.org/3/a0310e/A0310E00.htm#TOC.

Ballas, D. *et al.* (2007) 'Building a spatial microsimulation-based planning support system for local policy making', *Environment and Planning A*, 39, pp. 2482–2499. doi: 10.1068/a38441.

Bashir, S. A. (2002) 'Home is where the harm is', *American Journal of Public Health*, 92(5), pp. 733–8. doi: 10.2105/AJPH.92.5.733.

Bekele, S. (2017) 'The accuracy of demographic data in the Ethiopian Censuses', *Eastern Africa Social Science Research Review*, 33(1), pp. 15–38. doi: 10.1353/eas.2017.0001.

BIDS (2012) *Report of the post enumeration check (PEC) of the [Bangaldesh] Population and Housing Census, 2011*. Dhaka Bangladesh.

Biemer, P. P. (2010) 'Total survey error: Design, implementation, and evaluation', *Public Opinion Quarterly*, 74(5), pp. 817–848. doi: 10.1093/poq/nfq058.

Birkin, M. and Clarke, M. (1988) 'A synthetic spatial information system for urban and regional analysis: Methods and examples', *Environment and Planning A*, 20, pp. 1645–1671. doi: 10.1068/a201645.

Birkin, M. and Clarke, M. (1989) 'The generation of individual and household incomes at the small area level using synthesis', *Regional Studies*, 23(6), pp. 535–548. doi: 10.1080/00343408912331345702.

Boerma, J. T. and Sommerfelt, E. (1993) 'Demographic and Health Surveys (DHS): Contributions and limitations', *Rapp Trimest Statistic Sanit Mond*, 46, pp. 222–226.

Breiman, L. (1996) 'Bagging predictors', *Machine Learning*, 24, pp. 123–140. doi: 10.1023/A:1018054314350.

Breiman, L. (2001) 'Random forests', *Machine Learning*, 45, pp. 5–32. doi: 10.1023/A:1010933404324.

Brelsford, C. *et al.* (2018) 'Toward cities without slums: Topology and the spatial evolution of neighborhoods', *Science Advances*, 4(8), pp. 1–9. doi: 10.1126/sciadv.aar4644.

References

Brelsford, C., Martin, T. and Bettencourt, L. M. (2017) 'Optimal reblocking as a practical tool for neighborhood development', *Environment and Planning B*, 46(2), pp. 303–21. doi: 10.1177/2399808317712715.

Brown, V. J. (2003) 'The problem of inadequate or nonexistent housing', *Environmental Health Perspectives*, 111(2), pp. 92–99. Available at: https://ehp.niehs.nih.gov/doi/pdf/10.1289/ehp.111-a92.

Buhaug, H. and Urdal, H. (2013) 'An urbanization bomb? Population growth and social disorder in cities', *Global Environmental Change*. Elsevier Ltd, 23(1), pp. 1–10. doi: 10.1016/j.gloenvcha.2012.10.016.

Burgert, C. R., Zachary, B. and Colston, J. (2013) *Incorporating geographic information into demographic and health surveys: A field guide to GPS data collection*. Calverton, MD.

Bygbjerg, I. C. (2012) 'Double burden of noncommunicable and infectious diseases in developing countries', *Science*, 4, pp. 1499–1502. doi: 10.1126/science.1223466.

Cajka, J. *et al.* (2018) 'Geo-sampling in developing nations', *International Journal of Social Research Methodology*. Routledge, 21(6), pp. 729–746. doi: 10.1080/13645579.2018.1484989.

Campbell, L. (2017) *Working with people and communities in urban humanitarian crises*. London UK. Available at: https://www.alnap.org/system/files/content/resource/files/main/alnap-urban-people-and-communities-2017.pdf.

Campbell, T. and Campbell, A. (2007) 'Emerging disease burdens and the poor in cities of the developing world', *Journal of Urban Health*, 84(Suppl 1), pp. 54–64. doi: 10.1007/s11524-007-9181-7.

Carr-Hill, R. (2013) 'Missing millions and measuring development progress', *World Development*, 46, pp. 30–44. doi: 10.1016/j.worlddev.2012.12.017.

Carr-hill, R. O. Y. (2017) 'Improving population and poverty estimates with citizen surveys: Evidence from East Africa', *World Development*, 93, pp. 249–259. doi: 10.1016/j.worlddev.2016.12.017.

CDC (2014) *About BRFSS*, *Centers for Disease Control and Prevention (CDC)*. Available at: https://www.cdc.gov/brfss/about/index.htm (Accessed: 13 January 2020).

Center for International Earth Science Information Network (CIESIN) and Columbia University (2016) *Gridded Population of the World v4*. Available at:

http://sedac.ciesin.columbia.edu/data/collection/gpw-v4/sets/browse (Accessed: 2 February 2017).

Chapuis, K. *et al.* (2018) 'Gen*: A generic toolkit to generate spatially explicit synthetic populations', *International Journal of Geographical Information Science*. Taylor & Francis, 32(6), pp. 1–17. doi: 10.1080/13658816.2018.1440563.

Chen, S. *et al.* (2013) *Towards a post-2015 framework that counts: Developing national statistical capacity*. Paris France.

Chen, S. and Kalton, G. (2015) 'Geographic oversampling for race/ethnicity using data from the 2010 U.S. population census', *Journal of Survey Statistics and Methodology*, 3(4), pp. 543–565. doi: 10.1093/jssam/smv023.

Chew, R. F. *et al.* (2018) 'Residential scene classification for gridded population sampling in developing countries using deep convolutional neural networks on satellite imagery', *International Journal of Health Geographics*. BioMed Central, 17(1), pp. 1–17. doi: 10.1186/s12942-018-0132-1.

CIESIN *et al.* (2018) *Geo-Referenced Infrastructure and Demographic Data for Development (GRID3)*. Available at: https://www.grid3.org/ (Accessed: 13 January 2020).

CIESIN (2018) *Gridded Population of the World, Version 4.11 (GPWv4.11)*, *Center for International Earth Science Information Network - CIESIN - Columbia University*. Available at: https://doi.org/10.7927/H4F47M65 (Accessed: 27 October 2020).

Cobham, A. (2014) 'Uncounted: Power, inequalities and the post-2015 data revolution', *Development*, 57(3–4), pp. 320–337. doi: 10.1057/dev.2015.28.

Coghlan, B. *et al.* (2006) 'Mortality in the Democratic Republic of Congo: A nationwide survey', *Lancet*, 367, pp. 44–51.

Corsi, D. J. *et al.* (2012) 'Demographic and health surveys : a profile', *International Journal of Epidemiology*, 41, pp. 1602–1613. doi: 10.1093/ije/dys184.

CSO (2013) *[Zambia] 2010 Census of Population and Housing Post Enumeration Survey (PES)*. Lusaka Zambia. Available at: http://www.census.gov.ph/content/population-mountain-province-showed-increase14-thousand-results-2010-census-population-and.

Cutts, F. T. *et al.* (2016) 'Monitoring vaccination coverage: Defining the role of surveys', *Vaccine*, 34, pp. 4103–4109. doi: 10.1016/j.vaccine.2016.06.053.

References

Dabalen, A. *et al.* (2016) *Mobile phone panel surveys in developing countries: A practical guide for microdata collection*. Washington DC USA. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/24595/9781464809040.pdf?sequence=2.

Development Initiatives (2017) *Household surveys factsheet*. Bristol UK. Available at: http://devinit.org/wp-content/uploads/2017/07/Key-facts-on-household-surveys.pdf.

Development Initiatives (2018) *Global humanitarian assistance report 2018*. Bristol UK, Nairobi Kenya. doi: 10.1017/CBO9781107415324.004.

Deville, P. *et al.* (2014) 'Dynamic population mapping using mobile phone data', *Proceedings of the National Academy of Sciences*, 111(45), pp. 15888–15893. doi: 10.1073/pnas.1408439111.

DHS Program (2019) *Indicator data API*. Available at: http://api.dhsprogram.com (Accessed: 13 January 2020).

Diaz, M. M. *et al.* (2017) *State of world population 2017*. Edited by A. Erksen. New York NY USA: UN Population Fund (UNFPA). Available at: http://www.unfpa.org/swop.

DLR Earth Observation Center (2017) *Global Urban Footprint (GUF)*. Available at: http://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-11725/20508_read-47944/ (Accessed: 13 January 2020).

Dobson, J. E. *et al.* (2000) 'LandScan: A global population database for estimating populations at risk', *Photogrammetric Engineering and Remote Sensing*, 66(7), pp. 849–857.

Dooley, C. (2019) *gridEZ*, *Github*. Available at: https://github.com/cadooley/gridEZ (Accessed: 13 January 2020).

Doxsey-Whitfield, E. *et al.* (2015) 'Taking advantage of the improved availability of census data: A first look at the Gridded Population of the World, Version 4', *Papers in Applied Geography*. Routledge, 1(3), pp. 226–234. doi: 10.1080/23754931.2015.1014272.

Early, R. *et al.* (2016) 'Global threats from invasive alien species in the twenty-first century and national response capacities', *Nature communications*, 7. doi: 10.1038/ncomms12485.

Ebenstein, A. and Zhao, Y. (2015) 'Tracking rural-to-urban migration in China: Lessons from the 2005 inter-census population survey', *Population Studies*. Taylor & Francis, 69(3), pp. 337–353. doi: 10.1080/00324728.2015.1065342.

El-Zanaty, F. and Way, A. (2006) *Egypt Demographic and Health Survey 2005*. Cairo. Available at:

https://www.dhsprogram.com/pubs/pdf/FR176/FR176.pdf.

Elsey, H. *et al.* (2016) ‘Addressing inequities in urban health: Do decision-makers have the data they need? Report from the Urban Health Data special session at International Conference on Urban Health Dhaka 2015’, *Journal of Urban Health*. Journal of Urban Health, 93(3), pp. 526–537. doi: 10.1007/s11524-016-0046-9.

Elsey, H. *et al.* (2018) ‘Improving household surveys and use of data to address health inequities in three Asian cities: protocol for the Surveys for Urban Equity (SUE) mixed methods and feasibility study’, *BMJ Open*, 8(11), p. e024182. doi: 10.1136/bmjopen-2018-024182.

ESA-CCI (2017a) *Land Cover CCI Product - Annual LC maps from 2000 to 2015 (v2.0.7)*, *European Space Agency - Climate Change Initiative (ESA-CCI)*. Available at: http://maps.elie.ucl.ac.be/CCI/viewer/ (Accessed: 13 January 2020).

ESA-CCI (2017b) *Land cover CCI product - MERIS Waterbody product v4.0 (150 m)*, *European Space Agency - Climate Change Initiative (ESA-CCI)*. Available at: http://maps.elie.ucl.ac.be/CCI/viewer/ (Accessed: 13 January 2020).

ESRI (2018) ‘ArcGIS release 10’. Redlands CA USA: Environmental Systems Research Institute (ESRI).

European Commission (2008) *35 years of Eurobarometer*. Brussels Belgium. Available at: https://ec.europa.eu/commfrontoffice/publicopinion/docs/35_years_en.pdf.

European Commission (2011) *GEOSTAT 2011 grid dataset*. Available at: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography (Accessed: 13 January 2020).

European Commission (2012) *Representing census data in a European population grid (2010-2011)*. Brussels Belgium. Available at: http://ec.europa.eu/eurostat/documents/4311134/4350174/ESSnet-project-GEOSTAT1A-final-report_0.pdf.

European Commission (2017) *Global human settlement city model (GHS-SMOD)*. Available at: http://ghsl.jrc.ec.europa.eu/faq.php (Accessed: 13 January 2020).

European Commission JRC (2020) *Global human settlement population model (GHS-POP)*. Available at: https://ghsl.jrc.ec.europa.eu/data.php (Accessed: 27 October 2020).

European Space Agency (2012) *GlobCover*. Available at: www.esa-landcover-cci.org/?q=node/158

(Accessed: 13 January 2020).

Eurostat (2018) *EU Labour Force Survey development and history*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_force_survey_–_development_and_history#Development_of_the_EU-LFS (Accessed: 13 January 2020).

Ezeh, A. *et al.* (2017) 'The history, geography, and sociology of slums and the health problems of people who live in slums', *The Lancet*, 389, pp. 547–558. doi: 10.1016/S0140-6736(16)31650-6.

Fabic, M. S., Choi, Y. and Bird, S. (2012) 'A systematic review of Demographic and Health Surveys: Data availability and utilization for research', *Bulletin of the World Health Organization*, 90(8), pp. 604–612. doi: 10.2471/BLT.11.095513.

Facebook Connectivity Lab and CIESIN (2016) *High Resolution Settlement Layer (HRSL)*. Available at: https://data.humdata.org/dataset/highresolutionpopulationdensitymaps (Accessed: 27 October 2020).

Farrell, N., Morrissey, K. and O'Donoghue, C. (2012) 'Creating a spatial microsimulation model of the Irish local economy', in Tanton, R. and Edwards, K. (eds) *Spatial microsimulation: A reference guide for users*. Dordrecht: Springer, pp. 105–125.

de Ferranti, J. (2017a) *Digital elevation data: SRTM void fill - Viewfinder panoramas*. Available at: www.viewfinderPanoramas.org/voidfill.html (Accessed: 13 January 2020).

de Ferranti, J. (2017b) *Digital elevation data - Viewfinder panoramas*. Available at: http://www.viewfinderpanoramas.org/dem3.html (Accessed: 13 January 2020).

Fick, S. E. and Hijmans, R. J. (2017) 'WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas', *International Journal of Climatology*, 37(12), pp. 4302–4315. doi: 10.1002/joc.5086.

Flowminder Foundation (2019a) *GridSample.org*. Available at: www.gridsample.org (Accessed: 13 January 2020).

Flowminder Foundation (2019b) *GridSample2.0*, *GitHub*. Available at: https://github.com/Flowminder/GridSample2.0 (Accessed: 13 January 2020).

Frye, C. *et al.* (2018) 'Using classified and unclassified land cover data to estimate the footprint of human settlement', *Data Science Journal*, 17, pp. 1–12. doi: 10.5334/dsj-2018-020.

GADM (2015) *Global administrative areas version 3.6*. Available at: http://www.gadm.org (Accessed: 13 January 2020).

Gale, N. K. *et al.* (2013) 'Using the framework method for the analysis of qualitative data in multi-disciplinary health research', *BMC Medical Research Methodology*, 13(117), pp. 1–8. doi: 10.1186/1471-2288-13-117.

Gallup (2017a) *World Poll*. Available at: http://analytics.gallup.com/213704/world-poll.aspx (Accessed: 13 January 2020).

Gallup (2017b) *World Poll methodology*. Available at: https://news.gallup.com/poll/105226/world-poll-methodology.aspx (Accessed: 13 January 2020).

Galway, L. P. *et al.* (2012) 'A two-stage cluster sampling method using gridded population data, a GIS, and Google Earth imagery in a population-based mortality survey in Iraq', *International Journal of Health Geographics*, 11, p. 12. doi: 10.1186/1476-072X-11-12.

Gaughan, A. E. *et al.* (2013) 'High resolution population distribution maps for Southeast Asia in 2010 and 2015', *PLoS ONE*, 8(2), p. e55882. doi: 10.1371/journal.pone.0055882.

Gaughan, A. E. *et al.* (2015) 'Exploring nationally and regionally defined models for large area population mapping', *International Journal of Digital Earth*, 8(12), pp. 989–1006. doi: 10.1080/17538947.2014.965761.

GeoMarvel (2017) *GeoODK*. Available at: http://geoodk.com/ (Accessed: 9 June 2017).

Gibson, D. G. *et al.* (2017) 'Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review', *Journal of Medical Internet Research*, 19(5), pp. 1–17. doi: 10.2196/jmir.7428.

Gidado, S. O. *et al.* (2013) 'Polio field census and vaccination of underserved populations Northern Nigeria, 2012-2013', *Morbidity and Mortality Weekly Report*, 62(33), pp. 653–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23965827.

Gill, G. J. (2003) *Seasonal labour migration in rural Nepal: A preliminary overview*. 218. London UK. Available at: https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/2454.pdf.

Google LLC (2019) *Google Earth Pro v7.3.2.5776*. Available at: http://www.google.co.uk/earth/download/gep/agree.html (Accessed: 12 October 2019).

Grais, R. F., Rose, A. M. C. and Guthmann, J. (2007) 'Don't spin the pen: Two alternative methods for second-stage sampling in urban cluster surveys', *Emerging Themes in Epidemiology*, 7, pp. 1–7. doi: 10.1186/1742-7622-4-8.

GridSample (2019) *Case studies*. Available at: http://gridsample.org/case-studies (Accessed: 13 January 2020).

Grosh, M. E. and Glewwe, P. (1995) *A guide to Living Standards Measurement Study surveys and their data sets*. 120. Washington DC USA.

Grosh, M. G. E. and Munoz, J. (1996) *A manual for planning and implementing the Living Standards Measurement Study survey.* 126. Washington DC USA. Available at: http://documents.worldbank.org/curated/en/363321467990016291/pdf/multi-page.pdf.

Grosh, M. and Glewwe, P. (2000) *Designing household survey questionnaires for developing countries*. Washington DC USA. Available at: http://siteresources.worldbank.org/INTPOVRES/Resources/477227-1142020443961/2311843-1197996479165/part1_DesigningHHS.pdf.

Groves, R. M. *et al.* (2009) *Survey methodology*. 2nd edn. Hoboken NJ USA: Hohn Wiley & Sons, Inc.

Groves, R. M. (2011) 'Three eras of survey research', *The Public Opinion Quarterly*, 75(5), pp. 861–871. Available at: http://www.jstor.org/stable/41345914.

Groves, R. M. and Lyberg, L. (2010) 'Total survey error: Past, present, and future', *Public Opinion Quarterly*, 74(5), pp. 849–879. doi: 10.1093/poq/nfq065.

GSS (2012) *[Ghana] 2010 Population and Housing Census Post Enumeration Survey Report*. Accra Ghana. Available at: http/www.statsghana.gov.gh/../2010phc/010_Population_and Housing.

Gurgel, R. Q. *et al.* (2004) 'Capture-recapture to estimate the number of street children in a city in Brazil', *Archives of Disease in Childhood*, 89, pp. 222–224. doi: 10.1136/adc.2002.023481.

Haagsma, J. A. *et al.* (2016) 'The global burden of injury: Incidence, mortality, disability-adjusted life years and time trends from the global burden of disease study 2013', *Injury Prevention*, 22(1), pp. 3–18. doi: 10.1136/injuryprev-2015-041616.

Hagopian, A. *et al.* (2013) 'Mortality in Iraq associated with the 2003-2011 war and occupation: Findings from a national cluster sample survey by the University Collaborative Iraq Mortality Study', *PLoS Medicine*, 10(10), p. e1001533. doi: 10.1371/journal.pmed.1001533.

Hartigan, J. A. and Wong, M. A. (1979) 'Algorithm AS 136: A k-means clustering algorithm', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), pp. 100–108. Available at: www.jstor.org/stable/2346830.

Hay, S. *et al.* (2005) 'The accuracy of human population maps for public health application', *Tropical medicine & international health*, 10, pp. 1073–1086. doi: 10.1111/j.1365-3156.2005.01487.x.

HDX (2017) *Data*, *Humanitarian Data Exchange v.1.8.3*. Available at: https://data.humdata.org/ (Accessed: 13 January 2020).

Henderson, R. H. and Sundaresan, T. (1982) 'Cluster sampling to assess immunization coverage: A review of experience with a simplified sampling method', *Bulletin of the World Health Organization*, 60(2), pp. 253–260.

Heppenstall, A., Malleson, N. and Crooks, A. (2016) '"Space, the final frontier": How good are agent-based models at simulating individuals and space in cities?', *Systems*, 4(1), p. 9. doi: 10.3390/systems4010009.

Hillson, R. *et al.* (2014) 'Methods for determining the uncertainty of population estimates derived from satellite imagery and limited survey data: A case study of Bo City, Sierra Leone', *PLoS ONE*, 9(11). doi: 10.1371/journal.pone.0112241.

IAEG-SDGs (2017) *SDG indicators: Revised list of global Sustainable Development Goal indicators*, *Inter-Agency and Expert Group on Sustainable Development Goal Indicators*. Available at: https://unstats.un.org/sdgs/indicators/indicators-list/ (Accessed: 13 January 2020).

ICF International (2012a) *Demographic and Health Survey sampling and household listing manual*. Calverton MD USA. Available at: https://dhsprogram.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf.

ICF International (2012b) *Interviewer's manual*. Calverton MD USA. Available at: https://www.dhsprogram.com/pubs/pdf/DHSM1/DHS6_Interviewer_Manual_19Oct2012_DHSM1.pdf.

ICF International (2017) *Demographic and Health Surveys: Who we are*. Available at: http://dhsprogram.com/Who-We-Are/About-Us.cfm (Accessed: 13 January 2020).

ICF International (2020) *Available datasets*. Available at: https://dhsprogram.com/data/available-datasets.cfm (Accessed: 13 January 2020).

ICF Macro (2009) *Training field staff for DHS surveys*. Calverton MD USA. Available at: https://dhsprogram.com/pubs/pdf/DHSM3/Training_Field_Staff_for_DHS_Surveys_Oct2009.pdf.

IHME (2020) *Global Health Data Exchange (GHDx)*, *Institute for Health Metrics and Evaluation*

(IHME). Available at: http://ghdx.healthdata.org/ (Accessed: 13 January 2020).

IIPS and ICF International (2017) *National Family Health Survey (NFHS-4) 2015-16: India*. Mumbai India. Available at: https://dhsprogram.com/pubs/pdf/FR339/FR339.pdf.

IIPS and Macro International (2007) *National Family Health Survey (NFHS-3), 2005–06: India*. Mumbai India. Available at: https://dhsprogram.com/pubs/pdf/FRIND3/FRIND3-Vol1AndVol2.pdf.

INS and ORC Macro (1999) *Enquête Démographique et de Santé, Côte d'Ivoire 1998-1999*. Calverton MD USA.

International Statistical Institute (1973) 'The World Fertility Survey: An international programme of fertility research', *International Statistical Review*, 41(3), pp. 303–314. doi: 10.2307/1402619.

International Telecommunications Union (2018) *Statistics: Mobile-cellular subscriptions*. Available at: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx (Accessed: 13 January 2020).

IOM (2015) *Migration in Namibia*. Geneva Switzerland. Available at: https://publications.iom.int/system/files/pdf/mp_namibia_for_web_14june2016.pdf.

IRIN (1996) *Zaire: IRIN situation report on Idjwi Island, South Kivu, 8/6/96*, *Integrated Regional Information Network (IRIN)*. Available at: http://www.africa.upenn.edu/Hornet/irin_8696.html (Accessed: 13 January 2020).

Jennings, L. *et al.* (2015) 'Disparities in mobile phone access and maternal health service utilization in Nigeria: A population-based survey', *International Journal of Medical Informatics*. Elsevier Ireland Ltd, 84(5), pp. 341–348. doi: 10.1016/j.ijmedinf.2015.01.016.

Jiang, Q., Li, X. and Sánchez-barricarte, J. J. (2015) 'Data Uncertainties in China's Population', *Asian Social Science*, 11(13), pp. 200–205. doi: 10.5539/ass.v11n13p200.

Karanja, I. (2010) 'An enumeration and mapping of informal settlements in Kisumu, Kenya, implemented by their inhabitants', *Environment and Urbanization*, 22(1), pp. 217–239. doi: 10.1177/0956247809362642.

Katyal, A. *et al.* (2013) 'Using the Indian National Sample Survey data in public health research', *The National Medical Journal of India*, 26(5), pp. 291–294. Available at: https://www.researchgate.net/publication/263547309_Using_the_Indian_National_Sample_Survey_data_in_public_health_research.

Kauermann, G. and Küchenhoff, H. (2011) *Methoden und praktische Umsetzung mit R (Methods and practical implementation with R)*. Berlin/Heidelberg Germany: Springer-Verlag. doi: 10.1007 /

978-3-642-12318-4.

Kermorvant, C. *et al.* (2019) 'Spatially balanced sampling designs for environmental surveys', *Environmental Monitoring and Assessment*, 191(8). doi: 10.1007/s10661-019-7666-y.

Kish, L. (1965) *Survey sampling*. New York NY USA: Wiley.

KNBS *et al.* (2015) *Kenya Demographic and Health Survey 2014*. Rockville MD USA. Available at: https://dhsprogram.com/pubs/pdf/FR308/FR308.pdf.

Kools, L. (2018) *One-stage versus two-stage cluster sampling, a simulation study. In: Essays on wealth, health, and data collection.* Leiden University. Available at: https://openaccess.leidenuniv.nl/bitstream/handle/1887/67120/fulltext.pdf?sequence=1.

Korale, R. B. M. (2002) *Post Enumeration Survey 2001 [Nepal Population Census] Draft Report*. Kathmandu Nepal. Available at: http://202.45.144.7/cbsgov/nada/index.php/catalog/42/download/546.

Korn, E. L. and Graubard, B. I. (1998) 'Confidence intervals for proportions with small expected number of positive counts estimated from survey data', *Survey Methodology*, pp. 193–201. Available at: https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1998002/article/4356-eng.pdf?st=47IJQj9m.

Kronenfeld, D. A. (2008) 'Afghan refugees in Pakistan: Not all refugees, not always in Pakistan, not necessarily Afghan?', *Journal of Refugee Studies*, 21(1), pp. 43–63. doi: 10.1093/jrs/fem048.

Kuffer, M. *et al.* (2020) 'The role of earth observation in an integrated deprived area mapping "system" for low-to-middle income countries', *Remote Sensing*, 12(6). doi: 10.3390/rs12060982.

Kuffer, M., Pfeffer, K. and Sliuzas, R. (2016) 'Slums from Space—15 Years of Slum Mapping Using Remote Sensing', *Remote Sensing*, 8, pp. 1–29. doi: 10.3390/rs8060455.

Lai, S. *et al.* (2019) 'Exploring the use of mobile phone data for national migration statistics', *Palgrave Communications*, 5(1). doi: 10.1057/s41599-019-0242-9.

Lavely, W. (2001) 'First impressions from the 2000 Census of China', *Population and Development Review*, 27(4), pp. 755–769. doi: 10.1111/j.1728-4457.2001.00755.x.

Lê, T. A. and Verma, V. K. (1997) *An analysis of sample designs and sampling errors of the Demographic and Health Surveys*, *DHS Analytical Reports*. Calverton MD USA. Available at: https://dhsprogram.com/publications/publication-ar3-analytical-studies.cfm.

Lehner, B., Verdin, K. and Jarvis, A. (2006) *HydroSHEDS technical documentation*. Washington DC USA. Available at: http://www.worldwildlife.org/freshwater/pubs/HydroSHEDS_TechDoc_v10.pdf.

Lewis, G. L. (1983) 'The Contraceptive Prevalence Survey Project: Content and status', *Population Index*, 49(2), pp. 189–198. Available at: https://www.jstor.org/stable/2737212.

Leyk, S. *et al.* (2019) 'Allocating people to pixels: A review of large-scale gridded population data products and their fitness for use', *Earth System Science Data Discussions*, 11, pp. 1385–1409. doi: 10.5194/essd-2019-82.

Liaw, A. and Wiener, M. (2002) 'Classification and regression by randomForest', *R News*, pp. 18–22. Available at: https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf.

Lilford, R. *et al.* (2019) 'Because space matters: Conceptual framework to help distinguish slum from non-slum urban areas', *BMJ Global Health*, 4(2), p. e001267. doi: 10.1136/bmjgh-2018-001267.

Lilford, R. J. *et al.* (2017) 'Improving the health and welfare of people who live in slums', *The Lancet*, 389, pp. 559–570. doi: 10.1016/S0140-6736(16)31848-7.

Linard, C. *et al.* (2010) 'A high resolution spatial population database of Somalia for disease risk mapping.', *International journal of health geographics*, 9, p. 45. doi: 10.1186/1476-072X-9-45.

Linard, C. *et al.* (2012) 'Population distribution, settlement patterns and accessibility across Africa in 2010', *PloS one*, 7(2), p. e31743. doi: 10.1371/journal.pone.0031743.

Linard, C., Gilbert, M. and Tatem, A. J. (2011) 'Assessing the use of global land cover data for guiding large area population distribution modelling', *GeoJournal*, 76(5), pp. 525–538. doi: 10.1007/s10708-010-9364-8.

Lisowska, B. (2016) *Household surveys: Do competing standards serve country needs?* Bristol UK. Available at: http://devinit.org/wp-content/uploads/2018/01/Household-surveys-Do-competing-standards-serve-county-needs.pdf.

Lloyd, C. T. *et al.* (2019) 'Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets', *Big Earth Data*. Taylor & Francis, 3(2), pp. 108–139. doi: 10.1080/20964471.2019.1625151.

Lloyd, C. T., Sorichetta, A. and Tatem, A. J. (2017) 'High resolution global gridded data for use in population studies', *Nature Scientific Data*, 4, pp. 1–17. doi: 10.1038/sdata.2017.1.

Lohr, S. L. (2009) *Sampling: Design and Analysis*. 2nd edn. Boston MA USA: Brooks/Cole.

Long, J. F. and McMillen, D. B. (1987) 'A survey of census bureau population projection methods', *Climatic Change*, 11, pp. 141–177. doi: 10.1007/BF00138799.

Lu, X. *et al.* (2016) 'Detecting climate adaptation with mobile network data in Bangladesh: Anomalies in communication, mobility and consumption patterns during cyclone Mahasen', *Climatic Change*. Climatic Change, 138, pp. 505–519. doi: 10.1007/s10584-016-1753-7.

Lucci, P., Bhatkal, T. and Khan, A. (2018) 'Are we underestimating urban poverty?', *World Development*, 103, pp. 297–310. doi: 10.1016/j.worlddev.2017.10.022.

Lumley, T. (2004) 'Analysis of complex survey samples', *Journal of Statistical Software*, 9, pp. 1–19. doi: 10.18637/jss.v009.i08.

Macal, C. M. (2016) 'Everything you need to know about agent-based modelling and simulation', *Journal of Simulation*, 10(2), pp. 144–156. doi: 10.1057/jos.2016.7.

Mahabir, R. *et al.* (2018) 'A Critical Review of High and Very High-Resolution Remote Sensing Approaches for Detecting and Mapping Slums: Trends, Challenges and Emerging Opportunities', *Urban Science*, 2(1), p. 8. doi: 10.3390/urbansci2010008.

Mahapatra, P. *et al.* (2007) 'Civil registration systems and vital statistics: Successes and missed opportunities', *Lancet*, 370, pp. 1653–1663. doi: 10.1016/S0140-6736(07)61308-7.

Maro, R. (2009) *Post Enumeration Survey Tanzania Experience*, *Workshop on the 2010 World programme on population and housing censuses: census evaluation and post enumeration surveys, for English-speaking African countries*. Available at: https://unstats.un.org/unsd/demographic/meetings/wshops/Ethiopia_14_Sept_09/Country_Presentations/Tanzania.ppt (Accessed: 13 January 2020).

Marshall, S. J. (2004) 'Developing countries face double burden of disease', *Bulletin of the World Health Organization*, 82(7), p. 556. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2622909/pdf/15500291.pdf.

Martin, D., Cockings, S. and Leung, S. (2010) 'Progress report: 24-hour gridded population models', *European Forum for Geostatistics Conference, 05 - 07 Oct*. Tallinn Estonia. Available at: https://eprints.soton.ac.uk/175321/.

Martin, D., Lloyd, C. and Shuttleworth, I. (2011) 'Evaluation of gridded population models using 2001 Northern Ireland census data', *Environment and Planning A*, 43(8), pp. 1965–1980. doi:

10.1068/a43485.

Maxar (2019) *Satellite Imagery*, *Image Library*. Available at: https://www.digitalglobe.com/products/satellite-imagery (Accessed: 13 January 2020).

Mckenzie, D. J. and Mistiaen, J. (2009) 'Surveying migrant households: A comparison of census-based, snowball and intercept point surveys', *Journal of the Royal Statistical Society*, 172(2), pp. 339–360. doi: 10.1111/j.1467-985X.2009.00584.x.

MdP and Macro International (2008) *Enquête Démographique et de Santé, République Démocratique du Congo 2007*. Calverton MD USA. Available at: https://dhsprogram.com/pubs/pdf/FR208/FR208.pdf.

Meraner, A., Gumprecht, D. and Kowarik, A. (2016) 'Weighting procedure of the Austrian microcensus using administrative data', *Austrian Journal of Statistics*, 45(3), pp. 3–14. doi: 10.17713/ajs.v45i3.120.

Miles, P. (2016) 'Why more people are choosing to live on canal boats', *Financial Times*, 2 September. Available at: https://www.ft.com/content/9f3a3b7e-6aa6-11e6-a0b1-d87a9fea034f.

Minnesota Population Center (2018) *Integrated public use microdata series, international: version 7.0 [dataset]*. Minneapolis MN USA. doi: 10.18128/D010.V7.0.

MINSALUD and Profamilia (2015) *Colombia Encuesta Nacional de Demografía y Salud 2015*. Bogata Colombia. Available at: https://dhsprogram.com/pubs/pdf/FR334/FR334.pdf.

MoH, BoS and ICF Macro (2010) *Samoa Demographic and Health Survey 2009*. Apia Samoa. Available at: https://dhsprogram.com/pubs/pdf/fr240/fr240.pdf.

MoHSS and ICF International (2014) *Namibia Demographic and Health Survey 2013*. Windhoek Namibia and Rockville MD USA. Available at: https://dhsprogram.com/pubs/pdf/FR298/FR298.pdf.

Morris, L. (2000) 'History and current status of Reproductive Health Surveys at CDC', *American Journal of Preventive Medicine*, 19(1), pp. 31–34. doi: 10.1016/S0749-3797(00)00176-8.

Muñoz, J. and Langeraar, W. (2013) *A census-independent sampling strategy for a household survey in Myanmar*. Santiago Chile. Available at: http://winegis.com/images/census-independent-GIS-based-sampling-strategy-for-household-surveys-plan-of-action removed.pdf.

Murray, C. J. L. (2007) 'Towards good practice for health statistics: Lessons from the Millennium Development Goal health indicators', *Lancet*, 369, pp. 862–873. doi: 10.1016/S0140-

6736(07)60415-2.

My.com (2020) *MAPS.ME*. Available at: http://maps.me/en/home (Accessed: 13 January 2020).

Neilsen (2017) *Consumer panels*. Available at: http://www.nielsen.com/id/en/solutions/measurement/consumer-panels.html (Accessed: 13 January 2020).

Nelson, A. (2008) *Travel time to major cities: A global map of accessibility*. Available at: https://forobs.jrc.ec.europa.eu/products/gam/ (Accessed: 13 January 2020).

Newaya, T. P. (2010) *Rapid urbanization and its influence on the growth of informal settlements in Windhoek, Namibia [dissertation]*. Cape Peninsula University of Technology. Available at: http://digitalknowledge.cput.ac.za/xmlui/handle/11189/340.

Nieves, J. J. *et al.* (2020) 'Annually modelling built-settlements between remotely-sensed observations using relative changes in subnational populations and lights at night', *Computers, Environment and Urban Systems*, 80(May 2019), p. 101444. doi: 10.1016/j.compenvurbsys.2019.101444.

NISR (2010) *Post enumeration survey report: Fourth Population and Housing Census, Rwanda, 2012*. Kigali Rwanda. Available at: http://www.statistics.gov.rw/publication/rphc4-post-enumeration-survey.

NOAA (2012) *VIIRS nighttime lights*. Available at: https://maps.ngdc.noaa.gov/viewers/VIIRS_DNB_nighttime_imagery/index.html (Accessed: 13 January 2020).

NOAA (2017) *Version 4 DMSP-OLS Nighttime Lights Time Series*, *National Oceanic and Atmospheric Administration*. Available at: https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html (Accessed: 13 January 2020).

NPC and ICF International (2014) *Nigeria Demographic and Health Survey 2013*. Rockville MD USA. Available at: https://dhsprogram.com/pubs/pdf/fr293/fr293.pdf.

NSA-NPC (2016) *Household Income and Expenditure Survey 2015/2016 Report*. Windhoek Namibia. Available at: https://nsa.org.na/microdata1/index.php/catalog/28/related-materials.

NSA (2011a) *2011 Census EA boundaries*, *Digital Namibia*. Available at: https://nsa.org.na/page/gis-data-requests/ (Accessed: 13 January 2020).

NSA (2011b) *Namibia 2011 Population and Housing Census main report*. Windhoek Namibia. Available at: http://www.nsa.org.na/files/downloads/Namibia 2011 Population and Housing Census Main Report.pdf.

NSA (2013) *Namibia 2011 Population and Housing Census [PUMS dataset] version 1.0*. Windhoek Namibia. Available at: https://nsa.org.na/microdata1/index.php/catalog/19.

NSA (2016) *Namibia Intercensal Demographic Survey 2016 Report*. Windhoek Namibia. Available at: https://nsa.org.na/microdata1/index.php/catalog/30/related-materials.

NSA (2018) *Namibia Labour Force Survey 2018 Report*. Windhoek Namibia. Available at: https://nsa.org.na/page/central-data-catalogue/.

NSC (2014) *Census of India 2011: Report on post enumeration survey*. New Delhi India. Available at: http://censusindia.gov.in/2011Census/pes/pes_highlights.html.

Nuissl, H. and Heinrichs, D. (2013) 'Slums: Perspectives on the definition, the appraisal and the management of an urban phenomenon', *Journal of the Geographical Society of Berlin*, 144(2), pp. 105–116. doi: 10.12854/erde-144-8.

Okolo, A. (1999) 'The Nigerian Census: Problems and prospects', *The American Statistician*, 53(4), pp. 321–325. doi: 10.2307/2686050.

Oliveira, L. C. de S. *et al.* (2003) *Censo Demográfico 2000 - Pesquisa de avaliação da cobertura da coleta*. Rio de Janeiro Brazil. Available at: https://biblioteca.ibge.gov.br/biblioteca-catalogo.html?id=21402&view=detalhes.

OpenStreetMap contributors (2000) *OpenStreetMap base data*. Available at: www.openstreetmap.org (Accessed: 13 January 2020).

ORNL (2017) *LandScan documentation*. Available at: http://web.ornl.gov/sci/landscan/landscan_documentation.shtml (Accessed: 13 January 2020).

OsmAnd BV (2010) *OsmAnd offline mobile maps & navigation*. Available at: http://osmand.net/ (Accessed: 13 January 2020).

Oya, C. (2015) 'Who counts? Challenges and biases in defining "households" in research on poverty', *Journal of Development Effectiveness*, 7(3), pp. 336–345. doi: 10.1080/19439342.2015.1068358.

Pape, U. and Wollburg, P. (2019) *Estimation of poverty in Somalia using innovative methodologies*. 191. Washington DC USA. Available at:

http://documents.worldbank.org/curated/en/509221549985694077/Estimation-of-Poverty-in-Somalia-Using-Innovative-Methodologies.

Parry, H. (2018) ‘Thinking inside the box: The $600-a-month shipping container village for young professional getting creative to avoid staggering rents in San Francisco’, *Daily Mail Online*, 18 October. Available at: http://www.dailymail.co.uk/news/article-3277963/The-600-month-shipping-container-village-young-professionals-getting-creative-avoid-staggering-rents-San-Francisco.html.

Patel, N. N. *et al.* (2017) ‘Improving large area population mapping using geotweet densities’, *Transactions in GIS*, 21(2), pp. 317–331. doi: 10.1111/tgis.12214.

Peressini, T., Mcdonald, L. and David, J. (2010) ‘Towards a strategy for counting the homeless’, in *Finding Home: Policy Options for Addressing Homeless in Canada.* Toronto Canada: University of Toronto, p. 24. Available at: www.homelesshub.ca/Finding Home.

Perez-Heydrich, C. *et al.* (2016) ‘Influence of Demographic and Health Survey point displacements on raster-based analyses’, *Spatial Demography*, 4(2), pp. 135–153. doi: 10.1007/s40980-015-0013-1.

Perkins, T. A. *et al.* (2016) ‘Model-based projections of Zika virus infections in childbearing women in the Americas’, *Nature Microbiology*, 1(9), pp. 1–7. doi: 10.1038/nmicrobiol.2016.126.

Pesaresi, M., Melchiorri, M., *et al.* (2016) *Atlas of the human planet 2016: Mapping human presence on earth with the Global Human Settlement Layer*. Ispra Italy: European Commission Joint Research Centre. doi: 10.2788/582834.

Pesaresi, M., Ehrlich, D., *et al.* (2016) *Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014*. Ispra Italy: European Commission Joint Research Centre. doi: 10.2788/253582.

Pope, C. *et al.* (2000) ‘Analysing qualitative data’, *BMJ*, 320, pp. 114–6. doi: 10.1136/bmj.320.7227.114.

POPGRID Data Collaborative (2018) *POPGRID*. Available at: https://www.popgrid.org/compare-data (Accessed: 20 January 2020).

Qader, S. H. *et al.* (2019) *A novel approach to the automatic designation of pre-defined census enumeration areas and population sampling frames: A case study in Somalia*, *Policy Research Working Paper 8972*. Washington DC USA. Available at: http://documents.worldbank.org/curated/en/810771565268760366/A-Novel-Approach-to-the-

References

Automatic-Designation-of-Predefined-Census-Enumeration-Areas-and-Population-Sampling-Frames-A-Case-Study-in-Somalia.

QSR International (2018) *NVIVO*. Available at: https://www.qsrinternational.com/nvivo/support-overview/downloads (Accessed: 13 January 2020).

Randall, S. *et al.* (2015) 'UN census "households" and local interpretations in Africa since independence', *SAGE Open*, 5(2), pp. 1–18. doi: 10.1177/2158244015589353.

Randall, S. (2015) 'Where have all the nomads gone? Fifty years of statistical and demographic invisibilities of African mobile pastoralists', *Pastoralism*, 5, pp. 1–22. doi: 10.1186/s13570-015-0042-9.

Randall, S. and Coast, E. (2016) 'The quality of demographic data on older Africans', *Demographic Research*, 34(5), pp. 143–175. doi: 10.4054/DemRes.2016.34.5.

Richardson, L. F. *et al.* (2017) *Synthetic populations and ecosystems of the world (SPEW)*, *Department of Statistics and Data Science, Carnegie Mellon University*. Available at: http://www.stat.cmu.edu/~spew/about/ (Accessed: 13 January 2020).

Riley, L. *et al.* (2016) 'The World Health Organization STEPwise approach to noncommunicable disease risk-factor surveillance: Methods, challenges, and opportunities', *American Journal of Public Health*, 106(1), pp. 74–78. doi: 10.2105/AJPH.2015.302962.

Rimal, B. *et al.* (2017) 'Monitoring urban growth and the Nepal earthquake 2015 for sustainability of Kathmandu Valley, Nepal', *Land*, 6(42), pp. 1–23. doi: 10.3390/land6020042.

Rohde, J. *et al.* (2008) '30 years after Alma-Ata: Has primary health care worked in countries?', *Lancet*, 372, pp. 950–961. doi: 10.1016/S0140-6736(08)61405-1.

Romanuik, A. (2011) 'Persistence of high fertility in tropical Africa: The case of the Democratic Republic of the Congo', *Population and Development Review*, 37(1), pp. 1–28. doi: 10.1093/cercor/12.10.1048.

Roser, M., Ritchie, H. and Ortiz-Ospina, E. (2016) *World population growth*, *Our World in Data*. Available at: https://ourworldindata.org/world-population-growth/ (Accessed: 13 January 2020).

RTI International (2016) *RTI U.S. synthetic household population™*. Available at: https://www.rti.org/impact/synthpop (Accessed: 13 January 2020).

Rubin, D. B. (1987) *Multiple imputation for nonresponse in surveys*. New York NY USA: John Wiley & Sons.

Sabry, S. (2010) 'How poverty is underestimated in Greater Cairo, Egypt', *Environment and Urbanization*, 22(2), pp. 523–541. doi: 10.1177/0956247810379823.

Satterthwaite, D. (2010) *Urban myths and the mis-use of data that underpin them*. No. 2010/28. Helsinki Finland. Available at: https://www.wider.unu.edu/sites/default/files/wp2010-28.pdf.

Schneider, A., Friedl, M. A. and Potere, D. (2009) 'A new map of global urban extent from MODIS satellite data', *Environmental Research Letters*, 4, pp. 1–11. doi: 10.1088/1748-9326/4/4/044003.

Schneider, A., Friedl, M. A. and Potere, D. (2010) 'Mapping global urban areas using MODIS 500-m data: New methods and datasets based on "urban ecoregions"', *Remote Sensing of Environment*, 114, pp. 1733–1746. doi: 10.1016/j.rse.2010.03.003.

Selvin, H. C. (1958) 'Durkheim's suicide and problems of empirical research', *American Journal of Sociology*, 63(6), pp. 607–19. Available at: www.jstor.org/stable/2772991.

Setel, P. W. *et al.* (2007) 'A scandal of invisibility: Making everyone count by counting', *Lancet*, 370, pp. 2003–2006. doi: 10.1016/S0140-6736(07)61307-5.

Singh, R. and Mangat, N. S. (1996) 'Sampling from mobile populations', in *Elements of Survey Sampling*. Dordrecht Netherlands: Springer, pp. 314–330. doi: 10.1007/978-94-017-1404-4_12.

Sinha, P. *et al.* (2019) 'Assessing the spatial sensitivity of a random forest model: Application in gridded population modeling', *Computers, Environment and Urban Systems*. Elsevier, 75, pp. 132–145. doi: 10.1016/j.compenvurbsys.2019.01.006.

Skaarhoj, K. (2018) *ISSP history*, *International Social Survey Programme*. Available at: http://www.issp.org/about-issp/history/ (Accessed: 13 January 2020).

Skinner, C. (2018) 'Issues and challenges in census taking', *Annual Review of Statistics and Its Application*, 5, pp. 49–63. doi: 10.1146/annurev-statistics-041715-033713.

Snyder, R. E. *et al.* (2014) 'A comparison of social and spatial determinants of health between formal and informal settlements in a large metropolitan setting in Brazil', *Journal of Urban Health*, 91(3), pp. 432–445. doi: 10.1007/s11524-013-9848-1.

Sollom, R. *et al.* (2011) 'Health and human rights in Chin State, Western Burma: A population-based assessment using multistaged household cluster sampling', *PLoS Medicine*, 8(2), p. e1001007. doi: 10.1371/journal.pmed.1001007.

SSA (2012) *[South Africa] Census 2011 Post-enumeration survey*. Pretoria South Africa. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/485/download/8289.

References

Stark, L. *et al.* (2017) 'Estimating the size of the homeless adolescent population across seven cities in Cambodia', *BMC Medical Research Methodology*, 17, pp. 1–8. doi: 10.1186/s12874-017-0293-9.

StataCorps LLC (2020) *Stata software*. Available at: https://www.stata.com/ (Accessed: 13 January 2020).

Steven W., R. *et al.* (2004) 'A continuous satellite-derived measure of global terrestrial primary production', *Bioscience*, 54(6), pp. 547–560. doi: 10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2.

Stevens, F. R. *et al.* (2015) 'Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data', *PLoS ONE*, 10(2), p. e0107042. doi: 10.1371/journal.pone.0107042.

Subbaraman, R. *et al.* (2012) 'Off the map: The health and social implications of being a non-notified slum in India', *Environment and Urbanization*, 24(4), pp. 643–663. doi: 10.1177/0956247812456356.

Tacoli, C. (2009) 'Crisis or adaptation? Migration and climate change in a context of high mobility', *Environment and Urbanization*, 21(2), pp. 513–525. doi: 10.1177/0956247809342182.

Tatem, A. J. *et al.* (2007) 'High resolution population maps for low income nations: Combining land cover and census in East Africa', *PloS one*, 2(12), p. e1298. doi: 10.1371/journal.pone.0001298.

Tatem, A. J. *et al.* (2014) 'Mapping for maternal and newborn health: The distributions of women of childbearing age, pregnancies and births', *International Journal of Health Geographics*, 13, pp. 1–11. doi: 10.1186/1476-072X-13-2.

Tatem, A. J. *et al.* (2016) 'The geography of imported malaria to non-endemic countries: A meta-analysis of nationally reported statistics', *Lancet Infectious Diseases*, 3099(16). doi: 10.1016/S1473-3099(16)30326-7.

Tatem, A. J. (2017) 'WorldPop, open data for spatial demography', *Scientific Data*, 4, p. 170004. doi: 10.1038/sdata.2017.4.

Templ, M. *et al.* (2017) 'Simulation of synthetic complex data: The R package simPop', *Journal of Statistical Software*, 79(10), pp. 1–38. doi: 10.18637/jss.v079.i10.

Thomson, D. R. *et al.* (2012) 'Modelling strategic interventions in a population with a total fertility rate of 8.3: A cross-sectional study of Idjwi Island, DRC', *BMC public health*, 12(1), p. 959. doi: 10.1186/1471-2458-12-959.

Thomson, D. R. *et al.* (2016) *gridsample: Tools for grid-based survey sampling design*, *R package version 0.1.2*. Available at: https://cran.r-project.org/package=gridsample (Accessed: 13 January 2020).

Thomson, D. R. *et al.* (2017) 'GridSample: An R package to generate household survey primary sampling units (PSUs) from gridded population data', *International Journal of Health Geographics*, 16(1), pp. 1–19. doi: 10.1186/s12942-017-0098-4.

Thomson, D. R. *et al.* (2018) *Planning team guide v1.7*, *Surveys for Urban Equity (SUE)*. Available at: https://medicinehealth.leeds.ac.uk/downloads/download/95/planning_team_guide (Accessed: 13 January 2020).

Thomson, D. R. *et al.* (2019) 'Extending Data for Urban Health Decision-Making: a Menu of New and Potential Neighborhood-Level Health Determinants Datasets in LMICs', *Journal of Urban Health*, pp. 1–23. doi: 10.1007/s11524-019-00363-3.

Thomson, Dana R. *et al.* (2020) 'Addressing Unintentional Exclusion of Vulnerable and Mobile Households in Traditional Surveys in Kathmandu, Dhaka and Hanoi: A Mixed Methods Feasibility Study', *Journal of Urban Health*. doi: 10.1007/s11524-020-00485-z.

Thomson, Dana R, Rhoda, D. A., *et al.* (2020) 'Gridded population survey sampling: A systematic scoping review of the field and strategic research agenda', *International Journal of Health Geographics*, 19, p. 34. doi: 10.1186/s12942-020-00230-4.

Thomson, Dana R, Kuffer, M., *et al.* (2020) 'Need for an Integrated Deprived Area "Slum" Mapping System (IDEAMAPS) in Low- and Middle-Income Countries (LMICs)', *Social Science*, 9(5), p. 80. doi: 10.3390/socsci9050080.

Thomson, D. R. D. R. *et al.* (2021) 'Addressing unintentional exclusion of vulnerable and mobile households in traditional surveys in Kathmandu, Dhaka, and Hanoi: A mixed-methods feasibility study', *Journal of Urban Health*, 98(1), pp. 111–129. doi: 10.1007/s11524-020-00485-z.

Thomson, D. R., Kools, L. and Jochem, W. C. (2018) 'Linking synthetic populations to household geolocations: A demonstration in Namibia', *Data*, 3(3), p. 30. doi: 10.3390/data3030030.

Thorndike, E. (1939) 'On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them', *The American Journal of Psychology*, 52(1), pp. 122–4. doi: 10.2307/1416673.

References

Tipple, G. and Speak, S. (2005) 'Definitions of homelessness in developing countries', *Habitat International*, 29(2), pp. 337–352. doi: 10.1016/j.habitatint.2003.11.002.

Treiman, D. J. *et al.* (2005) *Observations on the design and implementation of sample surveys in China*. CCPR-006-05. Los Angeles CA USA. Available at: https://www.semanticscholar.org/paper/Observations-on-the-Design-and-Implementation-of-in-Treiman-Mason/615f1059443a3be862bcfa3aa46253ef109ad4e8.

Troisi, C. L. *et al.* (2015) 'Enhanced methodologies to enumerate persons experiencing homelessness in a large urban area', *Evaluation Review*, 39(5), pp. 480–500. doi: 10.1177/0193841X15610191.

UBS (2005) *Post enumeration survey: 2002 Uganda populaiton and housing census*. Entebbe Uganda. Available at: https://unstats.un.org/unsd/censuskb20/KnowledgebaseArticle10682.aspx.

UN-DESA (2014) *World Urbanization Prospects: 2014 Revision*. New York NY USA. Available at: http://esa.un.org/unpd/wup/FinalReport/WUP2014-Report.pdf.

UN-DESA (2017) *World population prospects: The 2017 revision*. New York. Available at: https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_Volume-I_Comprehensive-Tables.pdf.

UN-DESA (2018) *Sustainable Development Goals*, *Sustain Development Knowledge Platform*. Available at: https://sustainabledevelopment.un.org/sdgs (Accessed: 13 January 2020).

UN-DESA (2019) *World Urbanization Prospects: The 2019 Revision*. Available at: https://population.un.org/wup/DataQuery/ (Accessed: 13 January 2020).

UN-Habitat (2003) *Slums of the World: The face of urban poverty in the new millennium?*, *Global Urban Observatory*. Nairobi Kenya. Available at: http://www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=1124.

UN-Habitat (2016) *Urbanization and development: Emerging futures. World cities report 2016*. Nairobi Kenya: United Nations Human Settlements Programme (UN-Habitat). Available at: http://wcr.unhabitat.org/wp-content/uploads/2017/02/WCR-2016-Full-Report.pdf.

UN-Habitat (2017) *Distinguishing slum from non-slum areas to identify occupants' issues*, *News*. Available at: https://unhabitat.org/distinguishing-slum-from-non-slum-areas-to-identify-occupants-issues/ (Accessed: 13 January 2020).

UN-Habitat PSUP (2016) *Slum almanac 2015-2016*. Nairobi Kenya. Available at: https://unhabitat.org/slum-almanac-2015-2016/.

UN-OCHA-ROSA (2001a) *Namibia education facilities*, *Humanitarian Data Exchange*. Available at: https://data.humdata.org/organization/ocha-rosa (Accessed: 13 January 2020).

UN-OCHA-ROSA (2001b) *Namibia health facilities*, *Humanitarian Data Exchange*. Available at: https://data.humdata.org/organization/ocha-rosa (Accessed: 13 January 2020).

UN (1982) *Non-sampling errors in household surveys: Sources, assessment and control*, *National Household Survey Capability Programme*. New York NY USA. Available at: https://unstats.un.org/unsd/publication/unint/DP_UN_INT_81_041_2.pdf.

UNEP-WCMS and IUCN (2016) *World database on protected areas & Global database on protected areas management effectiveness*, *UN Environment Programme-World Conservation Monitoring Centre & International Union for Conservation of Nature*. Available at: https://www.protectedplanet.net/ (Accessed: 13 January 2020).

UNEP (2006) *Environmental data explorer: Gridded population of the world*. Available at: http://geodata.grid.unep.ch/ (Accessed: 13 January 2020).

UNFPA (2019) *UNFPA Strategy for the 2020 Round of Population and Housing Censuses (2015-2024)*. New York NY USA. Available at: https://www.unfpa.org/sites/default/files/pub-pdf/Census_Strategy_Final_July.pdf.

UNICEF (2000) *MICS 1 Conducting the field work*. New York NY USA. Available at: https://mics.unicef.org/tools?round=mics1#data-collection.

UNICEF (2013) *MICS 5 tools: Manual for mapping and household listing*. New York NY USA. Available at: http://mics.unicef.org/tools?round=mics5.

UNICEF (2015) *The Multiple Indicator Cluster Surveys (MICS) 1995-2015: Monitoring the situation of children and women for 20 years*. New York NY USA. Available at: http://54.92.12.252/files?job=W1siZiIsIjIwMTUvMDkvMTQvMTcvNTUvMzcvNTI2LzIwMTUwOTEyX01JQ1MyMF9XRUIucGRmIl1d&sha=da0e0b8ac785c628.

UNICEF (2017) *MICS 6 Instructions for interviewers*. New York NY USA. Available at: https://mics.unicef.org/tools#data-collection.

UNICEF (2019) *Multiple Indicator Cluster Surveys (MICS)*. Available at: http://mics.unicef.org/surveys (Accessed: 13 January 2020).

UNSD (2005a) *Designing household survey samples: Practical guidelines.* 98. New York NY USA. Available at: https://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf.

UNSD (2005b) *Household sample surveys in developing and transition countries*. New York NY USA. Available at: https://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf.

UNSD (2008) *Principles and recommendations for population and housing censuses - revision 2*. New York NY USA. Available at: https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67Rev2-E.pdf.

UNSD (2010) *Report on the results of a survey on census methods used by countries in the 2010 census round*. UNSD/DSSB/1. New York NY USA. Available at: http://unstats.un.org/unsd/census2010.htm.

UNSD (2016) *Demographic yearbook 2015*. New York NY USA. Available at: https://unstats.un.org/unsd/demographic/products/dyb/dyb2015.htm.

UNSD (2017) *Principles and recommendations for population and housing censuses - revision 3*. New York NY USA. Available at: https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf.

UNSD (2019) *2020 world population and housing census programme*, *Census dates for all countries*. Available at: https://unstats.un.org/unsd/demographic-social/census/censusdates/ (Accessed: 13 January 2020).

USCB (2013) *Demobase gridded mapping*, *Haiti, Pakistan, Rwanda*. Washington DC USA. Available at: https://www.census.gov/population/international/data/mapping/demobase.html (Accessed: 13 January 2020).

Wardrop, N. A. *et al.* (2018) 'Spatially disaggregated population estimates in the absence of national population and housing census data', *Proceedings of the National Academy of Sciences*, p. 201715305. doi: 10.1073/pnas.1715305115.

Weber, E. M. *et al.* (2018) 'Remote sensing of environment census-independent population mapping in northern Nigeria', *Remote Sensing of Environment*, 204, pp. 786–798. doi: 10.1016/j.rse.2017.09.024.

Weiss, D. *et al.* (2018) 'A global map of travel time to cities to assess inequalities in accessibility in 2015', *Nature*, 553(7688), pp. 333–6. doi: 10.1038/nature25181.

WFP-VAM (2018) *Urban essential needs assessment in the five communes of Kimbanseke, Kinsenso, Makala, N'sele and Selembao (Kinshasa)*. Rome Italy. Available at:

https://www.wfp.org/publications/democratic-republic-congo-urban-essential-needs-assessment-five-communes-kimbanseke-kinsenso.

WFP (2020) *Food security analysis*. Available at: https://vam.wfp.org (Accessed: 13 January 2020).

Whiteford, H. A. *et al.* (2013) 'Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010', *The Lancet*, 382(9904), pp. 1575–1586. doi: 10.1016/S0140-6736(13)61611-6.

WHO (2002) *Appendices (KISH tables, vignettes, cards)*. Geneva Switzerland. Available at: http://www.who.int/healthinfo/survey/whslongversionappendices.pdf.

WHO (2012) *Water sanitation hygiene: Key terms*. Available at: https://www.who.int/water_sanitation_health/monitoring/jmp2012/key_terms/en/ (Accessed: 13 January 2020).

WHO (2018) *Vaccination Coverage Cluster Surveys: Reference manual*. Geneva Switzerland. Available at: https://www.who.int/immunization/documents/who_ivb_18.09/en/.

Wilson, R. *et al.* (2016) 'Rapid and near real time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake', *PLoS Currents*, (1), pp. 1–26. doi: 10.1371/currents.dis.d073fbece328e4c39087bc086d694b5c.Authors.

World Bank (2018) *Data: Living Standards Measurement Study (LSMS)*. Available at: http://microdata.worldbank.org/index.php/catalog/lsms (Accessed: 21 November 2018).

World Bank (2020) *Survey Solutions*. Available at: https://mysurvey.solutions/ (Accessed: 13 January 2020).

WorldPop (2016) *Africa 1km internal migration flows*. Available at: https://www.worldpop.org/geodata/summary?id=1281 (Accessed: 13 January 2020).

WorldPop (2018a) *WorldPop-Global Namibia covariates*. Available at: ftp://ftp.worldpop.org.uk/GIS/Covariates/Global_2000_2020/NAM/ (Accessed: 13 January 2020).

WorldPop (2018b) *WorldPop-Global Namibia zonal statistics*. Available at: ftp://ftp.worldpop.org.uk/GIS/ZonalStatistics/Global_2000_2020/NAM/ (Accessed: 13 January 2020).

WorldPop (2019) *WorldPop-Global data 2000-2020*. Available at: https://www.worldpop.org/geodata/listing?id=29 (Accessed: 13 January 2020).

References

Woud, A. van der, Jan, de V. and Hayami, A. (1990) ‘Introduction: The hierarchies, provisioning, and demographic patterns of cities’, in Woud, A. van der, Jan, de V., and Hayami, A. (eds) *Urbanization in History*. Oxford UK: Clarendon Press, pp. 1–19.

Wright, J. D. and Devine, J. A. (1992) ‘Counting the homeless: The Census Bureau’s “S-Night” in five U.S. cities’, *Evaluation Review*, 16(4), pp. 355–364. Available at: http://erx.sagepub.com/cgi/content/abstract/16/4/376.

Yin, S. (2007) *Objections surface over Nigerian census results*, *Population Reference Bureau*. Available at: http://www.prb.org/Publications/Articles/2007/ObjectionsOverNigerianCensus.aspx (Accessed: 13 January 2020).

Zhang, Q., Pandey, B. and Seto, K. C. (2016) ‘A Robust Method to Generate a Consistent Time Series From DMSP / OLS Nighttime Light Data’, *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), pp. 1–11. doi: 10.1109/TGRS.2016.2572724.